

PCT



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 9/26, 1/20, C07H 21/04</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/25417</b> <b>(43) International Publication Date:</b> 17 July 1997 (17.07.97)
<b>(21) International Application Number:</b> PCT/US97/00092 <b>(22) International Filing Date:</b> 10 January 1997 (10.01.97)  <b>(30) Priority Data:</b> 08/583,787 11 January 1996 (11.01.96) US 08/712,612 13 September 1996 (13.09.96) US  <b>(71) Applicant:</b> RECOMBINANT BIOCATALYSIS, INC. [US/US]; 505 Coast Boulevard South, LaJolla, CA 92037 (US).  <b>(72) Inventors:</b> BYLINA, Edward, J.; West Court, Apartment A-1, Andalusia, PA 19020 (US). SWANSON, Ronald, V.; Apartment A, 309 No. Leuron Street, Media, PA 19063 (US). MATHUR, Eric, J.; 2654 Galicia Way, Carlsbad, CA 92009 (US). LAM, David, E.; 1518 West 249th Street, Harbor City, CA 90710 (US).  <b>(74) Agents:</b> HERRON, Charles, J. et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068-1739 (US).		<b>(81) Designated States:</b> AU, CA, IL, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> GLYCOSIDASE ENZYMES		
<b>(57) Abstract</b>  A thermostable glycosidase enzymes derived from various <i>thermococcus</i> , <i>staphylothermus</i> and <i>pyrococcus</i> organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## GLYCOSIDASE ENZYMES

This application is a continuation-in-part of pending patent application 08/583,787 filed January 11, 1996.

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention has been putatively identified as glucosidases,  $\alpha$ -galactosidases,  $\beta$ -galactosidases,  $\beta$ -mannosidases,  $\beta$ -mannanases, endoglucanases, and pullulanases.

The glycosidic bond of  $\beta$ -galactosides can be cleaved by different classes of enzymes: (i) phospho- $\beta$ -galactosidases (EC3.2.1.85) are specific for a phosphorylated substrate generated via phosphoenolpyruvate phosphotransferase system (PTS)-dependent uptake; (ii) typical  $\beta$ -galactosidases (EC 3.2.1.23), represented by the *Escherichia coli* LacZ enzyme, which are relatively specific for  $\beta$ -galactosides; and (iii)  $\beta$ -glucosidases (EC 3.2.1.21) such as the enzymes of *Agrobacterium faecalis*, *Clostridium thermocellum*, *Pyrococcus furiosus* or *Sulfolobus solfataricus* (Day, A.G. and Withers, S.G., (1986) Purification and characterization of a  $\beta$ -glucosidase from *Alcaligenes faecalis*. Can. J. Biochem. Cell. Biol. 64, 914-922; Kengen, S.W.M., et al. (1993) Eur. J. Biochem., 213, 305-312; Ait, N., Cruzet, N. and Cattaneo, J.

(1982) Properties of  $\beta$ -glucosidase purified from *Clostridium thermocellum*. J. Gen. Microbiol. 128, 569-577; Grogan, D.W. (1991) Evidence that  $\beta$ -galactosidase of *Sulfolobus solfataricus* is only one of several activities of a thermostable  $\beta$ -D-glycosidase. Appl. Environ. Microbiol. 57, 1644-1649). Members of the latter group, although highly specific with respect to the  $\beta$ -anomeric configuration of the glycosidic linkage, often display a rather relaxed substrate specificity and hydrolyse  $\beta$ -glucosides as well as  $\beta$ -fucosides and  $\beta$ -galactosides.

Generally,  $\alpha$ -galactosidases are enzymes that catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose.

Generally,  $\beta$ -mannanases are enzymes that catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising mannose groups.  $\beta$ -mannosidases hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of  $\beta$ -1,4 linked mannose backbone with  $\alpha$ -1,6 linked galactose sidechains. The enzymes required for the degradation of guar are  $\beta$ -mannanase,  $\beta$ -mannosidase and  $\alpha$ -galactosidase.  $\beta$ -mannanase hydrolyses the mannose backbone internally and  $\beta$ -mannosidase hydrolyses non-reducing, terminal mannose residues.  $\alpha$ -galactosidase hydrolyses  $\alpha$ -linked galactose groups.

Galactomannan polysaccharides and the enzymes that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a

need exists for thermostable galactomannases that are active in extreme conditions associated with drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is merit to eliminating raffinose from raw beet sugar.  $\alpha$ -Galactosidase has also been used as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

$\beta$ -Galactosidases which are active and stable at high temperatures appear to be superior enzymes for the production of lactose-free dietary milk products (Chaplin, M.F. and Bucke, C. (1990) In: Enzyme Technology, pp. 159-160, Cambridge University Press, Cambridge, UK). Also, several studies have demonstrated the applicability of  $\beta$ -galactosidases to the enzymatic synthesis of oligosaccharides via transglycosylation reactions (Nilsson, K.G.I. (1988) Enzymatic synthesis of oligosaccharides. Trends Biotechnol. 6, 156-264; Cote, G.L. and Tao, B.Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. 7, 145-162). Despite the commercial potential, only a few  $\beta$ -galactosidases of thermophiles have been characterized so far. Two genes reported are  $\beta$ -galactoside-cleaving enzymes of the hyperthermophilic bacterium *Thermotoga maritima*, one of the most thermophilic organotrophic eubacteria described to date (Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. and Stetter, K.O. (1986) *T. maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C, Arch. Microbiol. 144, 324-333) one of the most thermophilic organotrophic

eubacteria described to date. The gene products have been identified as a  $\beta$ -galactosidase and a  $\beta$ -glucosidase.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme hydrolyzes  $\alpha$ -1,6-glucosidic linkages on these polymers. Starch degradation for the production of sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with  $\alpha$ -amylase, and the second stage, or saccharification stage, is performed by  $\beta$ -amylase with pullulanase added as a debranching enzyme, to obtain better yields.

Endoglucanases can be used in a variety of industrial applications. For instance, the endoglucanases of the present invention can hydrolyze the internal  $\beta$ -1,4-glycosidic bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, and in the fruit juice and brewing industry for the clarification and extraction of juices.

The polynucleotides and polypeptides of the present invention have been identified as glucosidases,  $\alpha$ -galactosidases,  $\beta$ -galactosidases,  $\beta$ -mannosidases,  $\beta$ -mannanases, endoglucanases, and pullulanases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with another aspect of the present invention there are provided isolated nucleic acid molecules encoding mature polypeptides expressed by the DNA contained in ATCC Deposit No. 97379.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing lactose to galactose and glucose for use in the food processing industry, the pharmaceutical industry, for example, to treat intolerance to lactose, as a diagnostic reporter molecule, in corn wet milling, in the fruit juice industry, in baking, in the textile industry and in the detergent industry.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes for hydrolyzing guar gum (a galactomannan polysaccharide) to remove non-reducing terminal mannose residues. Further polysaccharides such as galactomannan and the enzymes according to the invention that degrade them have a variety of applications. Guar gum is commonly used as a thickening agent in food and also is utilized in hydraulic fracturing in oil and gas recovery. Consequently, mannanases are industrially relevant for the degradation and modification of guar gums. Furthermore, a need exists for thermostable mannanases that are active in extreme conditions associated with drilling and well stimulation.

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes

comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

#### Brief Description of the Drawings

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of M11TL of the present invention. Sequencing was performed using a 378 automated DNA sequencer for all sequences of the present invention (Applied Biosystems, Inc.).

Figure 2 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V-33B/G.

Figure 3 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of F1-12G.

Figure 4 are illustrations of the full-length DNA and corresponding deduced amino acid sequence of 9N2-31B/G.

Figure 5 are illustrations of the full-length DNA and corresponding deduced amino acid sequence of MSB8-6G.

Figure 6 are illustrations of the full-length DNA and corresponding deduced amino acid sequence of AEDII12RA-18B/G.

Figure 7 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of GC74-22G.



Figure 8 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of VC1-7G1.

Figure 9 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of 37GP1.

Figure 10 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of 6GC2.

Figure 11 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of 6GP2.

Figure 12 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of 63GB1.

Figure 13 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V.

Figure 14 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of 6GP3.

#### Definitions

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; i.e., produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is

transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

#### Summary of the Invention

In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode for the mature enzymes having the deduced amino acid sequences of Figures 1-14 (SEQ ID NOS:15-28).

In accordance with another aspect of the present invention, there are provided isolated polynucleotides encoding the enzymes of the present invention. The deposited material is a mixture of genomic clones comprising DNA encoding an enzyme of the present invention. Each genomic clone comprising the respective DNA has been inserted into a pBluescript vector (Stratagene, La Jolla, CA). The deposit has been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA, on December 13, 1995 and assigned ATCC Deposit No. 97379.

The deposit(s) have been made under the terms of the Budapest Treaty on the International Recognition of the deposit of micro-organisms for purposes of patent procedure. The strains will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit be required under 35 U.S.C. §112. The sequences of the polynucleotides contained in the deposited materials, as well as the amino acid sequences of the polypeptides encoded thereby, are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

#### Detailed Description of the Invention

The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

M11TL is a new species of *Desulfurococcus* isolated from Diamond Pool in Yellowstone National Park. The organism grows optimally at 85-88°C, pH 7.0 in a low salt medium containing yeast extract, peptone, and gelatin as substrates with a N<sub>2</sub>/CO<sub>2</sub> gas phase.

OC1/4V is from the genus *Thermotoga*. The organism was isolated from Yellowstone National Park. It grows optimally at 75°C in a low salt medium with cellulose as a substrate and N<sub>2</sub> in gas phase.

*Pyrococcus furiosus* VC1 is from the genus *Pyrococcus*. VC1 was isolated from Vulcano, Italy. It grows optimally at 100°C in a high salt medium (marine) containing elemental sulfur, yeast extract, peptone and starch as substrates and N<sub>2</sub> in gas phase.

*Staphylothermus marinus* F1 is a from the genus *Staphylothermus*. F1 was isolated from Vulcano, Italy. It grows optimally at 85°C, pH 6.5 in high salt medium (marine) containing elemental sulfur and yeast extract as substrates and N<sub>2</sub> in gas phase.

*Thermococcus* 9N-2 is from the genus *Thermococcus* 9N-2 was isolated from diffuse vent fluid in the East Pacific Rise. It is a strict anaerobe that grows optimally at 87°C.

*Thermotoga maritima* MSB8 is from the genus *Thermotoga*, and was isolated from Vulcano, Italy. MSB8 grows optimally at 85°C, pH 6.5 in a high salt medium (marine) containing starch and yeast extract as substrates and N<sub>2</sub> in gas phase.

*Thermococcus alcaliphilus* AEDII12RA is from the genus *Thermococcus*. AEDII12RA grows optimally at 85°C, pH 9.5 in a high salt medium (marine) containing polysulfides and yeast extract as substrates and N<sub>2</sub> in gas phase.

*Thermococcus chitonophagus* GC74 is from the genus *Thermococcus*. GC74 grows optimally at 85°C, pH 6.0 in a high salt medium (marine) containing chitin, meat extract, elemental sulfur and yeast extract as substrates and N<sub>2</sub> in gas

phase. AEPII 1a grows optimally at 85°C at pH 6.5 in marine medium under anaerobic conditions. It has many substrates.

[Add descriptions of new organisms]

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "M11TL" (Figure 1 and SEQ ID NOS:1 and 15), "OC1/4V-33B/G" (Figure 2 and SEQ ID NOS:2 and 16), "F1-12G" (Figure 3 and SEQ ID NOS:3 and 17), "9N2-31B/G" (Figure 4 and SEQ ID NOS:4 and 18), "MSB8" (Figure 5 and SEQ ID NOS:5 and 19), "AEDII12RA-18B/G" (Figure 6 and SEQ ID NOS:6 and 20), "GC74-22G" (Figure 7 and SEQ ID NOS:7 and 21), "VC1-7G1" (Figure 8 and SEQ ID NOS:8 and 22), "37GP1" (Figure 9 and SEQ ID NOS:9 and 23), "6GC2" (Figure 10 and SEQ ID NOS:10 and 24), "6GP2" (Figure 11 and SEQ ID NOS:11 and 25), "AEPII 1a" (Figure 12 and SEQ ID NOS:12 and 26), "OC1/4V" (Figure 13 and SEQ ID NOS:13 and 27), and "6GP3" (Figure 14 and SEQ ID NOS:28).

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

Clone	Gene/Protein with Closest Homology	Protein Identity	Nucleic Acid Identity
M11TL-29G	<i>Sulfolobus sulfataricus</i> DSM 1616/P1, $\beta$ -galactosidase	51%	55%
OC1/4V-33B/G	<i>Caldocellum saccharolyticum</i> , $\beta$ -glucosidase	52%	57%
<i>Staphylothermus marinus</i> F1-12G	<i>Bacillus polymyxa</i> , $\beta$ -galactosidase	36%	48%

<i>Thermococcus</i> 9N2-31B/G	<i>Sulfolobus</i> <i>sulfataricus</i> ATCC 49255/MT4, $\beta$ - galactosidase	51%	50%
<i>Thermotoga</i> <i>maritima</i> MSB8- 6G	<i>Clostridium</i> <i>thermocellum</i> bglB	45%	53%
<i>Thermococcus</i> AEDII12RA-18B/G	<i>Bacillus polymyxa</i> , $\beta$ -galactosidase	34%	48%
<i>Thermococcus</i> <i>chitonophagus</i> GC74-22G	<i>Sulfolobus</i> <i>sulfataricus</i> ATCC 49255/MT4, $\beta$ - galactosidase	46%	54%
<i>Pyrococcus</i> <i>furius</i> VC1- 7G1	<i>Sulfolobus</i> <i>sulfataricus</i> /MT-4 $\beta$ -galactosidase	46.4%	52.5%
<i>Thermotoga</i> <i>maritima</i> $\alpha$ - galactosidase (6GC2)	<i>Pediococcus</i> <i>pentosaceus</i> $\alpha$ - galactosidase	49%	29%
<i>Thermotoga</i> <i>maritima</i> $\beta$ - mannanase (6GP2)	<i>Aspergillus</i> <i>aculeatus</i> mannanase	56%	37%
AEPII 1a $\beta$ - mannosidase (63GB1)	<i>Sulfolobus</i> <i>solfataricus</i> $\beta$ - galactosidase	78%	56%
OC1/4V endoglucanase (33GP1)	<i>Clostridium</i> <i>thermocellum</i> endo- 1,4- $\beta$ - endoglucanase	65%	43%
<i>Thermotoga</i> <i>maritima</i> pullulanase (6GP3)	<i>Caldocellum</i> <i>saccharolyticum</i> $\alpha$ - destrom 6 glucanohydrolase	72	53
<i>Bankia gouldi</i> mix Endoglucanase (37GP1)	None available		

The polynucleotides and enzymes of the present invention show homology to each other as shown in Table 2.

Table 2

Clone	Gene/Protein with Closest Homology	Protein Identity	Nucleic Acid Identity
<i>Staphylothermus marinus</i> F1-12G	<i>Thermococcus</i> AEDII12RA-18B/G, $\beta$ -galactosidase, glucosidase	55%	57%
<i>Thermococcus</i> 9N2-31B/G	<i>Thermococcus chitonophagus</i> GC74-22G-glucosidase	74%	66%
<i>Pyrococcus furiosus</i> VC1-7G1	<i>Pyrococcus furiosus</i> VC1-7B/G $\beta$ -galactosidase	46.4%	54%

All the clones identified in Tables 1 and 2 encode polypeptides which have  $\alpha$ -glycosidase or  $\beta$ -glycosidase activity.

This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provide substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS:1-8; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS:1-8. Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:9-16, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, Ausubel F.M. et al. (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that the polynucleotides of SEQ ID NOS:1-14 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NOS:1-14 (i.e., comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.0, 5.0 mM Na<sub>2</sub>EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/mL polyriboadenylic acid. Approximately  $2 \times 10^7$  cpm (specific activity  $4-9 \times 10^8$  cpm/ug) of <sup>32</sup>P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na<sub>2</sub>EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at T<sub>m</sub> 10°C for the oligonucleotide probe. The membrane is then exposed to autoradiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. Further, it is understood that a section of

a 100 bps sequence that is 95 bps in length has 95% identity with the 1090 bps sequence from which it is obtained. See J. Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety. Also, it is understood that a fragment of a 100 bps sequence that is 95 bps in length has 95% identity with the 100 bps sequence from which it is obtained.

As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.

"Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NOS:1-8). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention relates polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.



It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. For example, gene libraries can be generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions can be performed on these libraries to generate libraries in the pBluescript phagemid. Libraries are thus generated and excisions performed according to the protocols/methods hereinafter described.

The excision libraries are introduced into the *E. coli* strain BW14893 F'kan1A. Expression clones are then identified using a high temperature filter assay. Expression clones encoding several glucanases and several other glycosidases are identified and repurified. The polynucleotides, and enzymes encoded thereby, of the present invention, yield the activities as described above.

The coding sequences for the enzymes of the present invention were identified by screening the genomic DNAs prepared for the clones having glucosidase or galactosidase activity.

An example of such an assay is a high temperature filter assay wherein expression clones were identified by use of high temperature filter assays using buffer Z (see recipe below) containing 1 mg/ml of the substrate 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucopyranoside (XGLU) (Diagnostic Chemicals Limited or Sigma) after introducing an excision library into the *E. coli* strain BW14893 F'kan1A. Expression clones encoding XGLUases were identified and repurified from M11TL,

OC1/4V, *Pyrococcus furiosus* VC1, *Staphylothemus marinus* F1, *Thermococcus* 9N-2, *Thermotoga maritima* MSB8, *Thermococcus alcaliphilus* AEDII12RA, and *Thermococcus chitonophagus* GC74.

Z-buffer: (referenced in Miller, J.H. (1992) A Short Course in Bacterial Genetics, p. 445.)

per liter:

$\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$	16.1g
$\text{NaH}_2\text{PO}_4 \cdot 7\text{H}_2\text{O}$	5.5g
KCl	0.75g
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	0.246g
$\beta$ -mercaptoethanol	2.7ml
Adjust pH to 7.0	

#### High Temperature Filter Assay

- (1) The f factor f'kan (from *E. coli* strain CSH118) (1) was introduced into the pho-pnh-lac-strain BW14893 (2). BW13893 (2). The filamentous phage library was plated on the resulting strain, BW14893 F'kan. (Miller, J.H. (1992) A Short Course in Bacterial Genetics; Lee, K.S., Metcalf, et al., (1992) Evidence for two phosphonate degradative pathways in *Enterobacter Aerogenes*, J. Bacteriol., 174:2501-2510.
- (2) After growth on 100 mm LB plates containing 100  $\mu\text{g}/\text{ml}$  ampicillin, 80  $\mu\text{g}/\text{ml}$  nethicillin and 1mM IPTG, colony lifts were performed using Millipore HATF membrane filters.
- (3) The colonies transferred to the filters were lysed with chloroform vapor in 150 mm glass petri dishes.
- (4) The filters were transferred to 100 mm glass petri dishes containing a piece of Whatman 3MM filter paper saturated with buffer.
  - (a) when testing for galactosidase activity (XGALase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGAL (ChemBridge Corporation). After transferring filter bearing lysed colonies to

the glass petri dish, placed dish in oven at 80-85°C.

(b) when testing for glucosidase (XGLUase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGLU. After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.

- (5) 'Positives' were observed as blue spots on the filter membranes. Used the following filter rescue technique to retrieve plasmid from lysed positive colony. Used pasteur pipette (or glass capillary tube) to core blue spots on the filter membrane. Placed the small filter disk in an Eppendorf tube containing 20  $\mu$ l water. Incubated the Eppendorf tube at 75°C for 5 minutes followed by vortexing to elute plasmid DNA off filter. This DNA was transformed into electrocompetent *E. coli* cells DH10B for *Thermatoga maritima* MSB8-6G, *Staphylothermus marinus* F1-12G, *Thermococcus AEDII12RA-18B/G*, *Thermococcus chitonophagus* GC74-22G, M11T1 and OC1/4V. Electrocompetent BW14893 F'kan1A *E. coli* were used for *Thermococcus* 9N2-31B/G, and *Pyrococcus furiosus* VC1-7G1. ~~Repeated filter-lift assay on transformation~~ plates to identify 'positives'. Return transformation plates to 37°C incubator after filter lift to regenerate colonies. Inoculate 3 ml LB liquid containing 100  $\mu$ g/ml ampicillin with repurified positives and incubate at 37°C overnight. Isolate plasmid DNA from these cultures and sequence plasmid insert. In some instances where the plates used for the initial colony lifts contained non-confluent colonies, a specific colony corresponding to a blue spot on the filter could be identified on a regenerated plate and repurified directly, instead of using the filter rescue technique.

Another example of such an assay is a variation of the high temperature filter assay wherein colony-laden filters

are heat-killed at different temperatures (for example, 105°C for 20 minutes) to monitor thermostability. The 3MM paper is saturated with different buffers (i.e., 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 100 mM Tris-Cl (pH 9.5)) to determine enzyme activity under different buffer conditions.

A  $\beta$ -glucosidase assay may also be employed, wherein Glcp $\beta$ Np is used as an artificial substrate (aryl- $\beta$ -glucosidase). The increase in absorbance at 405 nm as a result of p-nitrophenol (pNp) liberation was followed on a Hitachi U-1100 spectrophotometer, equipped with a thermostatted cuvette holder. The assays may be performed at 80°C or 90°C in closed 1-ml quartz cuvette. A standard reaction mixture contains 150 mM trisodium substrate, pH 5.0 (at 80°C), and 0.95 mM pNp derivative pNp = 0.561 mM<sup>-1</sup> • cm<sup>-1</sup>). The reaction mixture is allowed to reach the desired temperature, after which the reaction is started by injecting an appropriate amount of enzyme (1.06 ml final volume).

1 U  $\beta$ -glucosidase activity is defined as that amount required to catalyze the formation of 1.0  $\mu$ mol pNp/min. D-cellobiose may also be used as a substrate.

An ONPG assay for  $\beta$ -galactosidase activity is described by Miller, J.H. (1992) A Short Course in Bacterial Genetics and Mill, J.H. (1992) Experiments in Molecular Genetics, the contents of which are hereby incorporated by reference in their entirety.

A quantitative fluorometric assay for  $\beta$ -galactosidase specific activity is described by : Youngman P., (1987) Plasmid Vectors for Recovering and Exploiting Tn917 Transpositions in Bacillus and other Gram-Positive Bacteria. In Plasmids: A Practical approach (ed. K. Hardy) pp 79-103. IRL Press, Oxford. A description of the procedure can be found in Miller (1992) p. 75-77, the contents of which are incorporated by reference herein in their entirety.

The polynucleotides of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and

synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS:1-8) or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-14 (SEQ ID NOS:1-14).

The polynucleotide which encodes for the mature enzyme of Figures 1-14 (SEQ ID NOS:15-28) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-14 (SEQ ID NOS:15-28). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-14 (SEQ ID NOS:15-28) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-14 (SEQ ID NOS:15-28). Such nucleotide variants include deletion

variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-14 (SEQ ID NOS:1-14). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used,

the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-14 (SEQ ID NOS:1-14).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS:1-14, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS:15-28 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-14 (SEQ ID NOS:15-28) as well as fragments, analogs and derivatives of such enzyme.

The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-14 (SEQ ID NOS:15-28) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog

includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzymes of Figures 1-14 (SEQ ID NOS:15-28) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.



The enzymes of the present invention include the enzymes of SEQ ID NOS:15-28 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS:9-16 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS:15-28 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS:9-16 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and

pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, the phage lambda P<sub>L</sub> promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Bacillus subtilis; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed

to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P<sub>R</sub>, P<sub>L</sub> and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a

bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such

promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK),  $\alpha$ -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include E. coli, Bacillus subtilis, Salmonella typhimurium and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature

protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

$\beta$ -galactosidase hydrolyzes lactose to galactose and glucose. Accordingly, the OC1/4V, 9N2-31B/G, AEDII12RA-18B/G and F1-12G enzymes may be employed in the food processing industry for the production of low lactose content milk and for the production of galactose or glucose from lactose contained in whey obtained in a large amount as a by-product in the production of cheese. Generally, it is desired that enzymes used in food processing, such as the aforementioned  $\beta$ -galactosidases, be stable at elevated temperatures to help prevent microbial contamination.

These enzymes may also be employed in the pharmaceutical industry. The enzymes are used to treat intolerance to lactose. In this case, a thermostable enzyme is desired, as well. Thermostable  $\beta$ -galactosidases also have uses in diagnostic applications, where they are employed as reporter molecules.

Glucosidases act on soluble cellooligosaccharides from the non-reducing end to give glucose as the sole product. Glucanases (endo- and exo-) act in the depolymerization of cellulose, generating more non-reducing ends (endo-glucanases, for instance, act on internal linkages yielding cellobiose, glucose and cellooligosaccharides as products).  $\beta$ -glucosidases are used in applications where glucose is the



desired product. Accordingly, M11TL, F1-12G, GC74-22G and MSB8-6G (and OC1/4V, VC1-7G1, 9N2-31B/G and AEDII12RA18B/G) may be employed in a wide variety of industrial applications, including in corn wet milling for the separation of starch and gluten, in the fruit industry for clarification and equipment maintenance, in baking for viscosity reduction, in the textile industry for the processing of blue jeans, and in the detergent industry as an additive. For these and other applications, thermostable enzymes are desirable.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1  $\mu$ g of plasmid or DNA fragment is used with about 2 units of enzyme in about 20  $\mu$ l of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50  $\mu$ g of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for

particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. et al., *Nucleic Acids Res.*, 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., *Id.*, p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., *Virology*, 52:456-457 (1973).

#### Example 1

##### Bacterial Expression and Purification of Glycosidase Enzymes

DNA encoding the enzymes of the present invention, SEQ ID NOS:1 through 8, were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the respective PQE vector listed beneath

the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

*Thermococcus AEDII12RA* -18B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTA ACTATGGTGAATGCTATGATTGTC

(SEQ ID NO:29)

3' CGGAAGATCTTCATAGCTCCGGAAGCCCATA (SEQ ID NO:30)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

*OC1/4V-33B/G*

5' CCGAGAATTCATTAAAGAGGAGAAATTA ACTATGATAAGAAGGTCCGATTTTCC

(SEQ ID NO:31)

3' CGGAAGATCTTTAAGATTTTAGAAATTCCTT (SEQ ID NO:32)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

*Thermococcus 9N2* - 31B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTA ACTATGCTACCAGAAGGCTTTCTC

(SEQ ID NO:33)

3' CGGAGGTACCTCACCCAAGTCCGAACTTCTC (SEQ ID NO:34)

Vector: pQE30; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

*Staphylothermus marinus F1* - 12G

5' CCGAGAATTCATTAAAGAGGAGAAATTA ACTATGATAAGGTTTCCTGATTAT

(SEQ ID NO:35)

3' CGGAAGATCTTTATTTCGAGGTTCTTTAATCC (SEQ ID NO:36)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

*Thermococcus chitonophagus GC74* - 22G

5' CCGAGAATTCATTCAATTAAGAGGAGAAATTA ACTATGCTTCCAGGAGAACTTTCTC

(SEQ ID NO:37)

3' CGGAGGATCCCCTACCCCTCCTCTAAGATCTC (SEQ ID NO:38)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' BamHI.

M11TL

5' AATAATCTAGAGCATGCAATTCCCCAAAGACTTCATGATAG (SEQ ID NO:39)

3' AATAAAAGCTTACTGGATCAGTGTAAGATGCT (SEQ ID NO:40)

Vector: pQE70; and contains the following restriction enzyme sites 5' SphI and 3' Hind III.

*Thermotoga maritima* MSB8-6G

5' CCGACAATTGATTAAAGAGGAGAAATTAAGTATGGAAAGGATCGATGAAATT  
(SEQ ID NO:41)

3' CGGAGGTACCTCATGGTTTGAATCTCTTCTC (SEQ ID NO:42)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

*Pyrococcus furiosus* VC1 - 7G1

5' CCGACAATTGATTAAAGAGGAGAAATTAAGTATGTTCCCTGAAAAGTTCCTT  
(SEQ ID NO:43)

3' CGGAGGTACCTCATCCCCTCAGCAATTCCTC (SEQ ID NO:44)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Kpn I.

*Bankia gouldi* endoglucanase (37GP1)

5' AATAAGGATCCGTTTAGCGACGCTCGC  
(SEQ ID NO:45)

3' AATAAAAGCTTCCGGTTGTACAGCGGTAATAGGC (SEQ ID NO:46)

Vector: pQE52; and contains the following restriction enzyme sites 5' Bam HI and 3' Hind III.

*Thermotoga maritima*  $\alpha$ -galactosidase (6GC2)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAAGTATGATCTGTGTGGAAATATTCGGAAAG  
(SEQ ID NO:47)

3' TCTATAAAGCTTTTCACTCTCTCACCCTCTTCGTAGAAG (SEQ ID NO:48)

Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

*Thermotoga maritima*  $\beta$ -mannanase (6GP2)

5' TTTATTCAATTGATTAAAGAGGAGAAATTAAGTATGGGGATTGGTGGCGACGAC

(SEQ ID NO:49)

3' TTTATTAAGCTTATCTTTTCATATTCACATACCTCC (SEQ ID NO:50)

Vector: pQET; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

*AEPII 1a*  $\beta$ -mannanase (63GB1)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAAGTATGCTACCAGAAGAGTTCCTATGGGGC

(SEQ ID NO:51)

3' TTTATTAAGCTTCTCATCAACGGCTATGGTCTTCATTTC (SEQ ID NO:52)

Vector: pQET; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

*OC1/4V* endoglucanase (33GP1)

5' AAAAAACAATTGAATTCATTAAAGAGGAGAAATTAAGTATGGTAGAAAGACACTTCAGATATGTTCTT

(SEQ ID NO:53)

3' TTTTTCGGATCCAATTCATTACTCTTTGCCTG (SEQ ID NO:54)

Vector: pQET; and contains the following restriction enzyme sites 5' BamHI and 3' EcoRI.

*Thermotoga maritima* pullulanase (6GP3)

5' TTTTGGGAATTCATTAAAGAGGAGAAATTAAGTATGGAACTGATCATAGAAGGTTAC

(SEQ ID NO:55)

3' ATAAGAAGCTTTTCACTCTCTGTACAGAACGTACGC (SEQ ID NO:56)

Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp'), a bacterial origin of replication (ori), an IPTG-regulatable

promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

#### Example 2

##### Isolation of A Selected Clone From the Deposited genomic clones

A clone is isolated directly by screening the deposited material using the oligonucleotide primers set forth in Example 1 for the particular gene desired to be isolated. The specific oligonucleotides are synthesized

using an Applied Biosystems DNA synthesizer. The oligonucleotides are labeled with  $^{32}\text{P}$ -ATP using T4 polynucleotide kinase and purified according to a standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The deposited clones in the pBluescript vectors may be employed to transform bacterial hosts which are then plated on 1.5% agar plates to the density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed DNA is prehybridized in 6 x SSC, 20 mM  $\text{NaH}_2\text{PO}_4$ , 0.4% SDS, 5 x Denhardt's 500  $\mu\text{g/ml}$  denatured, sonicated salmon sperm DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6xSSC, 20 mM  $\text{NaH}_2\text{PO}_4$ , 0.4% SDS, 500  $\mu\text{g/ml}$  denatured, sonicated salmon sperm DNA with  $1 \times 10^6$  cpm/ml  $^{32}\text{P}$ -probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones are isolated and purified by secondary and tertiary screening. The purified clone is sequenced to verify its identity to the primer sequence.

Once the clone is isolated, the two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25  $\mu\text{l}$  of reaction mixture with 0.5  $\mu\text{g}$  of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM  $\text{MgCl}_2$ , 0.01% (w/v) gelatin, 20  $\mu\text{M}$  each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the



DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product. The ends of the newly purified genes are nucleotide sequenced to identify full length sequences. Complete sequencing of full length genes is then performed by Exonuclease III digestion or primer walking.

### Example 3

#### Screening for Galactosidase Activity

Screening procedures for  $\alpha$ -galactosidase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Dilute XL1-Blue MRF *E. coli* host of (Stratagene Cloning Systems, La Jolla, CA) to O.D.<sub>600</sub> = 1.0 with NZY media. In 15 ml tubes, inoculate 200  $\mu$ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) containing 1mM IPTG to each tube and pour onto all NYZ plate surface. Allow to cool and incubate at 37 °C overnight. The assay plates are obtained as substrate p-Nitrophenyl  $\alpha$ -galactosidase (Sigma) (200 mg/100 ml) (100 mM NaCl, 100 mM Potassium-Phosphate) 1% (w/v) agarose. The plaques are overlayed with nitrocellulose and incubated at 4 °C for 30 minutes whereupon the nitrocellulose is removed and overlayed onto the substrate plates. The substrate plates are then incubated at 70 °C for 20 minutes.

### Example 4

#### Screening of Clones for Mannanase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for  $\beta$ -mannanase activity.

A culture solution of the Y1090-*E. coli* host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to

O.D.<sub>600</sub>=1.0 with NZY media. The amplified library from *Thermotoga maritima* lambda gtl1 library was diluted in SM (phage dilution buffer):  $5 \times 10^7$  pfu/ $\mu$ l diluted 1:1000 then 1:100 to  $5 \times 10^2$  pfu/ $\mu$ l. Then 8  $\mu$ l of phage dilution ( $5 \times 10^2$  pfu/ $\mu$ l) was plated in 200  $\mu$ l host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UV™ nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

An Azo-galactomannan overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% Azocarob-galactomannan. (Megazyme, Australia). The plates were incubated at 72 °C. The Azocarob-galactomannan treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the Azocarob-galactomannan plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500  $\mu$ l SM (phage dilution buffer) and 25  $\mu$ l CHCl<sub>3</sub>.

#### Example 5

#### Screening of Clones for Mannosidase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for  $\beta$ -mannosidase activity.

A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.<sub>600</sub>=1.0 with NZY media. The amplified library from AEPII 1a lambda gt11 library was diluted in SM (phage dilution buffer):  $5 \times 10^7$  pfu/ $\mu$ l diluted 1:1000 then 1:100 to  $5 \times 10^2$  pfu/ $\mu$ l. Then 8  $\mu$ l of phage dilution ( $5 \times 10^2$  pfu/ $\mu$ l) was plated in 200  $\mu$ l host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UV™ nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

A p-nitrophenyl- $\beta$ -D-manno-pyranoside overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% p-nitrophenyl- $\beta$ -D-manno-pyranoside. (Megazyme, Australia). The plates were incubated at 72 °C. The p-nitrophenyl- $\beta$ -D-manno-pyranoside treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the p-nitrophenyl- $\beta$ -D-manno-pyranoside plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking.

the individual portions in 500  $\mu$ l SM (phage dilution buffer) and 25  $\mu$ l CHCl<sub>3</sub>.

#### Example 6

##### Screening for Pullulanase Activity

Screening procedures for pullulanase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Host cells are diluted to O.D.<sub>600</sub> = 1.0 with NZY or appropriate media. In 15 ml tubes, inoculate 200  $\mu$ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) is added to each tube and the mixture is plated, allowed to cool, and incubated at 37°C for about 28 hours. Overlays of 4.5 mls of the following substrate are poured:

##### 100 ml total volume

0.5g	Red Pullulan Red (Megazyme, Australia)
1.0g	Agarose
5ml	Buffer (Tris-HCL pH 7.2 @ 75 °C)
2ml	5M NaCl
5ml	CaCl <sub>2</sub> (100mM)
85ml	dH <sub>2</sub> O

Plates are cooled at room temperature, and then incubated at 75°C for 2 hours. Positives are observed as showing substrate degradation.

#### Example 7

##### Screening for Endoglucanase Activity

Screening procedures for endoglucanase protein activity may be assayed for as follows:

1. The gene library is plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates (~4,800 plaque forming units/plate) in E.coli host with LB agarose as top agarose. The plates are incubated at 37°C overnight.

2. Plates are chilled at 4°C for one hour.
3. The plates are overlayed with Duralon membranes (Stratagene) at room temperature for one hour and the membranes are oriented and lifted off the plates and stored at 4°C.
4. The top agarose layer is removed and plates are incubated at 37°C for ~3 hours.
5. The plate surface is rinsed with NaCl.
6. The plate is stained with 0.1% Congo Red for 15 minutes.
7. The plate is destained with 1M NaCl.
8. The putative positives identified on plate are isolated from the Duralon membrane (positives are identified by clearing zones around clones). The phage is eluted from the membrane by incubating in 500µl SM + 25µl CHCl<sub>3</sub> to elute.
9. Insert DNA is subcloned into any appropriate cloning vector and subclones are reassayed for CMCase activity using the following protocol:
  - i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.
  - ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.
  - iii) Incubate at 37°C for 2 hours.
  - iv) Stain with 0.1% Congo Red for 15 minutes.
  - v) Destain with 1M NaCl for 15 minutes.
  - vi) Identify positives by clearing zone around clone.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a member selected from the group consisting of:
  - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding an enzyme comprising amino acid sequences set forth in SEQ ID NOS:15-28;
  - (b) a polynucleotide which is complementary to the polynucleotide of (a); and
  - (c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) or (b).
2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 which encodes an enzyme comprising an amino acid sequence which a member selected from the group
  - (a) according to SEQ ID NO:15;
  - (b) according to SEQ ID NO:16;
  - (c) according to SEQ ID NO:17;
  - (d) according to SEQ ID NO:18;
  - (e) according to SEQ ID NO:19;
  - (f) according to SEQ ID NO:20;
  - (g) according to SEQ ID NO:21;
  - (h) according to SEQ ID NO:22;
  - (i) according to SEQ ID NO:23;
  - (j) according to SEQ ID NO:24;
  - (k) according to SEQ ID NO:25;
  - (l) according to SEQ ID NO:26;
  - (m) according to SEQ ID NO:27; and
  - (n) according to SEQ ID NO:28.

5. An isolated polynucleotide comprising a member selected from the group consisting of:
- (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding an enzyme encoded by the DNA contained in ATCC Deposit No. 97379, wherein said enzyme is selected from the group consisting of M11TL, OC1/4V, F1-12G, 9N2-31B/G, MSB8-6G, AEDII12RA-18B/G, GC74-22G and VC1-7G1;
  - (b) a polynucleotide complementary to the polynucleotide of (a); and
  - (c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) and (b).
6. A vector comprising the DNA of Claim 2.
7. A host cell comprising the vector of Claim 6.
8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.
- 
9. A process for producing a cell comprising: transforming or transfecting the cell with the vector of Claim 6 such that the cell expresses the polypeptide encoded by the DNA contained in the vector.
10. An enzyme comprising a member selected from the group consisting of:
- (a) an enzyme comprising an amino acid sequence which is at least 70% identical to the amino acid sequence set forth in SEQ ID NOS:15-28; and
  - (b) an enzyme which comprises at least 30 amino acid residues to the enzyme of (a).

11. A method for generating glucose from soluble cellooligosaccharides comprising:  
administering an effective amount of an enzyme selected from the group consisting of an enzyme having the amino acid sequence set forth in SEQ ID NOS:15-28.



M11TL GLYCOSIDASE - 29G  
COMPLETE GENE SEQUENCE - 9/95

1	TTG AAA TTT GGT AAA GAG TTT ATG ATA GAG TAC TCA TTT TCA GCG TTT CAA TTT GAA GAG	60
2	Met Lys Phe Pro Lys Asp Phe Met Ile Gly Tyr Ser Ser Ser Pro Phe Gln Phe Gln Ala	20
61	GCT ATT GGT GGT TGT CAG GAT GGT AAT ACT GAT TGG TGG CTA TGG GTC CAT GAT GGT GAG	120
71	Gly Ile Pro Gly Ser Gln Asp Pro Asn Ser Asp Trp Trp Val Trp Val His Asp Pro Gln	40
121	AAC ACA GCA GGT GCA CTA CTC AGT GGT GAT TTT GGT CAG AAC GGC CCA GGT TAT TGG AAT	180
131	Asn Thr Ala Ala Gly Leu Val Ser Gly Asp Phe Pro Gln Asn Gly Pro Gly Tyr Trp Asn	60
181	TTA AAC CAA AAT GAC CAC GAG CTC GGT GAG AAG CTG GCG GTT AAC ACT ATT AGA GTA GGC	240
191	Leu Asn Gln Asn Asp His Asp Leu Ala Glu Lys Leu Gly Val Asn Thr Ile Arg Val Gly	80
241	GTT GAG TGG AGT AGG ATT TTT CCA AAG CCA ACT TTC AAT GTT AAA GTC CCT GTA GAG AGA	300
251	Val Glu Trp Ser Arg Ile Phe Pro Lys Pro Thr Phe Asn Val Lys Val Pro Val Glu Arg	100
301	GAT GAG AAC GGC AGC ATT GTT CAC GTA GAT GTC GAT GAT AAA GCG GTT GAA AGA CTT GAT	360
311	Asp Glu Asn Gly Ser Ile Val His Val Asp Val Asp Asp Lys Ala Val Glu Arg Leu Asp	120
361	GAA TTA GCC AAC AAG GAG GCC GTA AAC CAT TAC GTA GAA ATG TAT AAA GAC TGG GTT GAA	420
371	Glu Leu Ala Asn Lys Glu Ala Val Asn His Tyr Val Glu Met Tyr Lys Asp Trp Val Glu	140
421	AGA GGT AGA AAA CTT ATA CTC AAT TTA TAC CAT TGG CCC CTG CCT CTC TGG CTT CAC AAC	480
431	Arg Gly Arg Lys Leu Ile Leu Asn Leu Tyr His Trp Pro Leu Pro Leu Trp Leu His Asn	160
481	CCA ATC ATG GTG AGA AGA ATG GGC CCG GAC AGA GCG CCC TCA GGC TGG CTT AAC GAG GAG	540
491	Pro Ile Met Val Arg Arg Met Gly Pro Asp Arg Ala Pro Ser Gly Trp Leu Asn Glu Glu	180
541	TCC GTG GTG GAG TTT GCC AAA TAC GCC GCA TAC ATT GCT TGG AAA ATG GGC GAG CTA CCT	600
551	Ser Val Val Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala Trp Lys Met Gly Glu Leu Pro	200
601	GTT ATG TGG AGC ACC ATG AAC GAA CCC AAC GTC GTT TAT GAG CAA GGA TAC ATG TTC GTT	660
611	Val Met Trp Ser Thr Met Asn Glu Pro Asn Val Val Tyr Glu Gln Gly Tyr Met Phe Val	220
661	AAA GGG GGT TTC CCA CCC GGC TAC TTG AGT TTG GAA GCT GCT GAT AAG GCC AGG AGA AAT	720
671	Lys Gly Gly Phe Pro Pro Gly Tyr Leu Ser Leu Glu Ala Ala Asp Lys Ala Arg Arg Asn	240
721	ATG ATC CAG GCT CAT GCA CGG GCC TAT GAC AAT ATT AAA CGC TTC AGT AAG AAA CCT GTT	780
731	Met Ile Gln Ala His Ala Arg Ala Tyr Asp Asn Ile Lys Arg Phe Ser Lys Lys Pro Val	260
781	GGA CTA ATA TAC GCT TTC CAA TGG TTC GAA CTA TTA GAG GGT CCA GCA GAA GTA TTT GAT	840
791	Gly Leu Ile Tyr Ala Phe Gln Trp Phe Glu Leu Leu Glu Gly Pro Ala Glu Val Phe Asp	280
841	AAG TTT AAG AGC TCT AAG TTA TAC TAT TTC ACA GAC ATA GTA TCG AAG GGT AGT TCA ATC	900
851	Lys Phe Lys Ser Ser Lys Leu Tyr Tyr Phe Thr Asp Ile Val Ser Lys Gly Ser Ser Ile	300
901	ATC AAT GTT GAA TAC AGG AGA GAT CTT GCC AAT AGG CTA GAC TGG TTG GGC GTT AAC TAC	960
911	Ile Asn Val Glu Tyr Arg Arg Asp Leu Ala Asn Arg Leu Asp Trp Leu Gly Val Asn Tyr	320
961	TAT AGC CGT TTA GTC TAC AAA ATC GTC GAT GAC AAA CCT ATA ATC CTG CAC GGC TAT GGA	1020
971	Tyr Ser Arg Leu Val Tyr Lys Ile Val Asp Asp Lys Pro Ile Ile Leu His Gly Tyr Gly	340
1021	TTC CTT TGT ACA CCT GGG GGG ATC AGC CCG GCT GAA AAT CCT TGT AGC GAT TTT GCG TGG	1080
1031	Phe Leu Cys Thr Pro Gly Gly Ile Ser Pro Ala Glu Asn Pro Cys Ser Asp Phe Gly Trp	360
1081	GAG GTC TAT CCT GAA GGA CTC TAC CTA CTT CTA AAA GAA CTT TAC AAC CGA TAC GGG GTA	1140
1091	Glu Val Tyr Pro Glu Gly Leu Tyr Leu Leu Leu Lys Glu Leu Tyr Asn Arg Tyr Gly Val	380
1141	GAC TTG ATC GTC ACT GAG AAC GGT GTT TCA GAC AGC AAG GAT GCG TTG AGA CCG GCA TAC	1200
1151	Asp Leu Ile Val Thr Glu Asn Gly Val Ser Asp Ser Arg Asp Ala Leu Arg Pro Ala Tyr	400
1201	CTG GTC TGG CAT GTT TAC AGC GTA TGG AAA GGC GCT AAC GAG GGC ATT CCC GTC AAA GGC	1260
1211	Leu Val Ser His Val Tyr Ser Val Trp Lys Ala Ala Asn Glu Gly Ile Pro Val Lys Gly	420
1261	TAC CTT GAC TGG AGT TTT ACA GAG AAT TAT GAG TGG GGT GAG GGC TTT AGC GAG AAA TTT	1320
1271	Tyr Leu His Trp Ser Leu Thr Asp Asn Tyr Glu Trp Ala Glu Gly Phe Arg Gln Lys Phe	440

Figure 1

1401 GAG GTA GTC ATG GTC CAG TTT AAA AGC AAG AAA AAG TAT GTC CCA CCA AAG GAG CTA GTC 1400  
441 GAG GAG VAL AGC VAL AGC TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
1401 TTT TTT GAG ATG GCA AGC CAG AAG CCA ATA CTC GAT GAG CTA CAG CAT CTT ACA CTC ATC 1400  
441 TTT AGC GTC TTT ATA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
1441 CAG TAA 1440  
481 Gln End 482

Figure 1 (Continued)

OC1/4 GLYCOSIDASE - 33G/B  
COMPLETE GENE SEQUENCE - 9/95

1	ATG ATA AGA AGG TCC GAT TTT CCA AAA GAT TTT ATC TTC GCA AGG GGT ACC GCA GCA TAC	60
1	Met Ile Arg Arg Ser Asp Phe Pro Lys Asp Phe Ile Phe Gly Thr Ala Thr Ala Ala Tyr	20
61	TAG ATT GAA GGT GCA GCA AAC GAA GAT GGC AGA GGG CCA TCA ATT TGG GAT GTC TTT TCA	120
21	Gln Ile Glu Gly Ala Ala Asn Glu Asp Gly Arg Gly Pro Ser Ile Trp Asp Val Phe Ser	40
121	CAC ACG CCT GGC AAA ACC CTG AAC GGT GAC ACA GGA GAC GTT GCG TGT GAC CAT TAT CAC	180
41	His Thr Pro Gly Lys Thr Leu Asn Gly Asp Thr Gly Asp Val Ala Cys Asp His Tyr His	60
181	CGA TAC AAG GAA GAT ATC CAG CTG ATG AAA GAA ATA GGG TTA GAC GCT TAC AGG TTC TCT	240
61	Arg Tyr Lys Glu Asp Ile Gln Leu Met Lys Glu Ile Gly Leu Asp Ala Tyr Arg Phe Ser	80
241	ATC TCC TGG CCC AGA ATT ATG CCA GAT GGG AAG AAC ATC AAC CAA AAG GGT GTG GAT TTC	300
81	Ile Ser Trp Pro Arg Ile Met Pro Asp Gly Lys Asn Ile Asn Gln Lys Gly Val Asp Phe	100
301	TAC AAC AGA CTC GTT GAT GAG CTT TTG AAG AAT GAT ATC ATA CCA TTC GTA ACA CTC TAT	360
101	Tyr Asn Arg Leu Val Asp Glu Leu Leu Lys Asn Asp Ile Ile Pro Phe Val Thr Leu Tyr	120
361	CAC TGG GAC TTA CCC TAC GCA CTT TAT GAA AAA GGT GGA TGG CTT AAC CCA GAT ATA GCG	420
121	His Trp Asp Leu Pro Tyr Ala Leu Tyr Glu Lys Gly Gly Trp Leu Asn Pro Asp Ile Ala	140
421	CTC TAT TTC AGA GCA TAC GCA ACG TTT ATG TTC AAC GAA CTC GGT GAT CGT GTG AAA CAT	480
141	Leu Tyr Phe Arg Ala Tyr Ala Thr Phe Met Phe Asn Glu Leu Gly Asp Arg Val Lys His	160
481	TGG ATT ACA CTG AAC GAA CCA TGG TGT TCT TCT TTC TCG GGT TAT TAC ACG GGA GAG CAT	540
161	Trp Ile Thr Leu Asn Glu Pro Trp Cys Ser Ser Phe Ser Gly Tyr Tyr Thr Gly Glu His	180
541	GCC CCG GGT CAT CAA AAT TTA CAA GAA GCG ATA ATC GCG GCG CAC AAC CTG TTG AGG GAA	600
181	Ala Pro Gly His Gln Asn Leu Gln Glu Ala Ile Ile Ala Ala His Asn Leu Leu Arg Glu	200
601	CAT GGA CAT GCC GTC CAG GCG TCC AGA GAA GAA GTA AAA GAT GGG GAA GTT GGC TTA ACC	660
201	His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Gly Leu Thr	220
661	AAC GTT GTG ATG AAA ATA GAA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT	720
221	Asn Val Val Met Lys Ile Glu Pro Gly Asp Ala Lys Pro Glu Ser Phe Leu Val Ala Ser	240
721	CTT GTT GAT AAG TTC GTT AAT GCA TGG TCC CAT GAC CCT GTT GTT TTC GGA AAA TAT CCC	780
241	Leu Val Asp Lys Phe Val Asn Ala Trp Ser His Asp Pro Val Val Phe Gly Lys Tyr Pro	260
781	GAA GAA GCA GTT GCA CTT TAT ACG GAA AAA GCG TTG CAA GTT CTC GAT AGC GAT ATG AAT	840
261	Glu Glu Ala Val Ala Leu Tyr Thr Glu Lys Gly Leu Gln Val Leu Asp Ser Asp Met Asn	280
841	ATT ATT TCG ACT CCT ATA GAC TTC TTT GGT GTG AAT TAT TAC ACA AGA ACA CTT GTT GTT	900
281	Ile Ile Ser Thr Pro Ile Asp Phe Phe Gly Val Asn Tyr Tyr Thr Arg Thr Leu Val Val	300
901	TTT GAT ATG AAC AAT CCT CTT GGA TTT TCG TAT GTT CAG GGA GAC CTT CCC AAA ACG GAG	960
301	Phe Asp Met Asn Asn Pro Leu Gly Phe Ser Tyr Val Gln Gly Asp Leu Pro Lys Thr Glu	320
961	ATG GGA TGG GAA ATC TAC CCG CAG GGA TTA TTT GAT ATG CTG GTC TAT CTG AAG GAA AGA	1020
321	Met Gly Trp Glu Ile Tyr Pro Gln Gly Leu Phe Asp Met Leu Val Tyr Leu Lys Glu Arg	340
1021	TAT AAA CTA CCA CTT TAT ATC ACA GAG AAC GGG ATG GCT GCA CCT GAT AAA TTG GAA AAC	1080
341	Tyr Lys Leu Pro Leu Tyr Ile Thr Glu Asn Gly Met Ala Gly Pro Asp Lys Leu Glu Asn	360
1081	GGA AGA GTT CAT GAT AAT TAC CGA ATT GAA TAT TTG GAA AAG CAC TTT GAA AAA GCA CTT	1140
361	Gly Arg Val His Asp Asn Tyr Arg Ile Glu Tyr Leu Glu Lys His Phe Glu Lys Ala Leu	380
1141	GAA GCA ATC AAT GCA GAT GTT GAT TTG AAA GGT TAC TTC ATT TGG TCT TTG ATG GAT AAC	1200
381	Glu Ala Ile Asn Ala Asp Val Asp Leu Lys Gly Tyr Phe Ile Trp Ser Leu Met Asp Asn	400
1201	TTC GAA TGG GCG TGC GGA TAC TCC AAA CGT TTC GGT ATA ATC TAC GTA GAT TAC AAT ACC	1260
401	Phe Glu Trp Ala Cys Gly Tyr Ser Lys Arg Phe Gly Ile Ile Tyr Val Asp Tyr Asn Thr	420
1261	CCA AAA AGG ATA TTT AAA GAT TCA GCG ATG TCG TTG AAG GAA TTT CTA AAA TCT TAA	1317
421	Pro Lys Arg Ile Leu Lys Asp Ser Ala Met Trp Leu Lys Glu Phe Leu Lys Ser End	419

Figure 2

STAPHYLOTHERMUS MARINUS GLYCOSIDASE - 12G  
COMPLETE GENE SEQUENCE  
9/95

1	TTG ATA AGG TTT CCT GAT TAT TTC TTG TTT GGA ACA GGT ACA TCA TCG CAC CAG ATT GAG	60
1	Met Ile Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile Gln	20
61	GGT AAT AAC ATA TTT AAT GAT TGG TGG GAG TGG GAG ACT AAA GGC AGG ATT AAG GTT AGA	120
21	Gly Asn Asn Ile Phe Asn Asp Trp Trp Glu Trp Glu Thr Lys Gly Arg Ile Lys Val Arg	40
121	TCG GGT AAG GCA TGT AAT CAT TGG GAA CTC TAT AAA GAA GAC ATA GAG CTT ATG GCT GAG	180
41	Ser Gly Lys Ala Cys Asn His Trp Glu Leu Tyr Lys Glu Asp Ile Glu Leu Met Ala Glu	60
181	CTG GGA TAT AAT GCT TAT AGG TTC TCC ATA GAG TGG AGT AGA ATA TTT CCC AGA AAA GAT	240
61	Leu Gly Tyr Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phe Pro Arg Lys Asp	80
241	CAT ATA GAT TAT GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TAC	300
81	His Ile Asp Tyr Glu Ser Leu Asn Lys Tyr Lys Glu Ile Val Asn Leu Leu Arg Lys Tyr	100
301	GGG ATA GAA CCT GTA ATC ACT CTT CAC CAC TTC ACA AAC CCG CAA TGG TTT ATG AAA ATT	360
101	Gly Ile Glu Pro Val Ile Thr Leu His His Phe Thr Asn Pro Gln Trp Phe Met Lys Ile	120
361	GGT GGA TCG ACT AGG GAA GAG AAC ATA AAA TAT TTT ATA AAA TAT GTA GAA CTT ATA GCT	420
121	Gly Gly Trp Thr Arg Glu Glu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Glu Leu Ile Ala	140
421	TCC GAG ATA AAA GAC GTG AAA ATA TGG ATC ACT ATT AAT GAA CCA ATA ATA TAT GTT TTA	480
141	Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asn Glu Pro Ile Ile Tyr Val Leu	160
481	CAA GGA TAT ATT TCC GGC GAA TGG CCA CCT GGA ATT AAA AAT TTA AAA ATA CCT GAT CAA	540
161	Gln Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asn Leu Lys Ile Ala Asp Gln	180
541	GTA ACT AAG AAT CTT TTA AAA GCA CAT AAT GAA GCC TAT AAT ATA CTT CAT AAA CAC GGT	600
181	Val Thr Lys Asn Leu Leu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly	200
601	ATT GTA GGC ATA GCT AAA AAC ATG ATA GCA TTT AAA CCA GGA TCT AAT AGA GGA AAA GAC	660
201	Ile Val Gly Ile Ala Lys Asn Met Ile Ala Phe Lys Pro Gly Ser Asn Arg Gly Lys Asp	220
661	ATT AAT ATT TAT CAT AAA GTC GAT AAA GCA TTC AAC TGG GGA TTT CTC AAC GGA ATA TTA	720
221	Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Leu Asn Gly Ile Leu	240
721	AGG GGA GAA CTA GAA ACT CTC CGT GGA AAA TAC CGA GTT GAG CCC GGA AAT ATT GAT TTC	780
241	Arg Gly Glu Leu Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Pro Gly Asn Ile Asp Phe	260
781	ATA GGC ATA AAC TAT TAT TCA TCA TAT ATT GTA AAA TAT ACT TGG AAT CCT TTT AAA CTA	840
261	Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr Thr Trp Asn Pro Phe Lys Leu	280
841	CAT ATT AAA GTC GAA CCA TTA GAT ACA GGT CTA TGG ACA ACT ATG GGT TAC TGC ATA TAT	900
281	His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Met Gly Tyr Cys Ile Tyr	300
901	CCT AGA GGA ATA TAT GAA GTT GTA ATG AAA ACT CAT GAG AAA TAC GGC AAA GAA ATA ATC	960
301	Pro Arg Gly Ile Tyr Glu Val Val Met Lys Thr His Glu Lys Tyr Gly Lys Glu Ile Ile	320
961	ATT ACA GAG AAC GGT GTT GCA GTA GAA AAT GAT GAA TTA AGG ATT TTA TCC ATT ATC AGG	1020
321	Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg	340
1021	CAC TTA CAA TAC TTA TAT AAA CCC ATG AAT GAA GGA GCA AAG GTG AAA GCA TAT TTC TAC	1080
341	His Leu Gln Tyr Leu Tyr Lys Ala Met Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr	360
1081	TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA	1140
361	Trp Ser Phe Met Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val	380
1141	GAA GTT GAT TAT AAG ACT TTT GAG ACA AAA CCT AGA AAA AGC GCA TAT GTA TAT AGT CAA	1200
381	Glu Val Asp Tyr Lys Thr Phe Glu Arg Lys Pro Arg Lys Ser Ala Tyr Val Tyr Ser Gln	400
1201	ATA GCA CGT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC	1260
401	Ile Ala Arg Thr Lys Thr Ile Ser Asp Glu Tyr Leu Glu Lys Tyr Gly Leu Lys Asn Leu	420
1261	GAA TAA	1266
421	Glu End	422

Figure 3

Thermococcus 9N2 Glycosidase 318/0  
Complete gene sequence 9/95

1	ATG CTA CCA GAA GGC TTT CTC TGG GGC GTG TCC CAG TCC GGC TTT CAG TTC GAG ATG GGC	60
1	Met Leu Pro Glu Gly Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Gln Phe Glu Met Gly	70
61	GAC AAG CTC AGG AGG AAC ATT GAT CCG AAC ACA GAC TCG TCG AAG TCG GTC AGG GAT CCC	120
21	Asp Lys Leu Arg Arg Asn Ile Asp Pro Asn Thr Asp Trp Trp Lys Trp Val Arg Asp Pro	40
121	TTC AAC ATA AAG AGG GAA CTC GTC AGC GGC GAC CTG CCC GAG GAG GGG ATA AAC AAC TAC	180
41	Phe Asn Ile Lys Arg Glu Leu Val Ser Gly Asp Leu Pro Glu Glu Gly Ile Asn Asn Tyr	60
181	GAA CTT TAC GAG AAG GAT CAC CGC CTC GGC AGA GAC CTC GGT CTG AAC GTT TAC AGG ATT	240
61	Glu Leu Tyr Glu Lys Asp His Arg Leu Ala Arg Asp Leu Gly Leu Asn Val Tyr Arg Ile	80
241	GGA ATA GAG TGG AGC AGG ATC TTT CCC TGG CCA ACG TGG TTT GTG GAG GTT GAC GTT GAG	300
81	Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp Pro Thr Trp Phe Val Glu Val Asp Val Glu	100
301	CGG GAC AGC TAC GGA CTC GTG AAG GAC GTC AAA ATC GAT AAA GAC ACG CTC GAA GAG CTC	360
101	Arg Asp Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Asp Lys Asp Thr Leu Glu Glu Leu	120
361	GAC GAG ATA GCG AAT CAT CAG GAG ATA CCC TAC TAC CGC CGC GTT ATA GAG CAC CTC AGC	420
121	Asp Glu Ile Ala Asn His Gln Glu Ile Ala Tyr Tyr Arg Arg Val Ile Glu His Leu Arg	140
421	GAG CTC GGC TTC AAG CTC ATC GTG AAC CTC AAC CAC TTC ACG CTC CCC CTC TCG CTT CAC	480
141	Glu Leu Gly Phe Lys Val Ile Val Asn Leu Asn His Phe Thr Leu Pro Leu Trp Leu His	160
481	GAT CCC ATA ATC GCG AGC GAG AAG GGC CTC ACC AAC GGT ACG ATT GGC TCG CTC GGC CAG	540
161	Asp Pro Ile Ile Ala Arg Glu Lys Ala Leu Thr Asn Gly Arg Ile Gly Trp Val Gly Gln	180
541	GAG ACC GTC GTG GAG TTC GGC AAG TAC GCG GCG TAC ATC GCG AAC GCA CTC GGG GAC CTC	600
181	Glu Ser Val Val Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala Asn Ala Leu Gly Asp Leu	200
601	CTT CAT ATG TGG AGC ACC TTC AAC GAG CCG ATG GTC GTT GTG GAG CTC GGT TAC CTC GCG	660
201	Val Asp Met Trp Ser Thr Phe Asn Glu Pro Met Val Val Val Glu Leu Gly Tyr Leu Ala	220
661	CCC TAC TCC GGC TTT CCG CCG GGG GTT ATG AAC CCC GAG GCG GCA AAG CTC GCA ATC CTC	720
221	Pro Tyr Ser Gly Phe Pro Pro Gly Val Met Asn Pro Glu Ala Ala Lys Leu Ala Ile Leu	240
721	AAC ATG ATA AAC GGC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC ACG GTA AAG	780
241	Asn Met Ile Asn Ala His Ala Leu Ala Tyr Lys Met Ile Lys Lys Phe Asp Arg Val Lys	260
781	GCC GAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GGC GTT	840
261	Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly Val	280
841	GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC TAC	900
281	Ala Tyr Pro Tyr Asp Ser Asn Asp Pro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn Tyr	300
901	TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GGC AAG CTC AAC ATC GAG TTC GAC	960
301	Phe His Ser Gly Leu Phe Phe Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe Asp	320
961	GCT GAG ACC TTC GTC AAA GTT CCG CAT CTC ACG GCG AAC GAC TCG ATA GGC GTT AAC TAC	1020
321	Gly Glu Thr Phe Val Lys Val Arg His Leu Arg Gly Asn Asp Trp Ile Gly Val Asn Tyr	340
1021	TAC ACG AGA GAA GTC GTC AGG TAT TCG GAG CCC AAG TTC CCG ACG ATA CCC CTC ATA TCC	1080
341	Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro Leu Ile Ser	360
1081	TTC CCG GCA GTT CAC AAC TAC GGC TAC GCC TGC AGG CCC GCG ACT TCT TCC GCC GAC GGA	1140
361	Phe Arg Gly Val His Asn Tyr Gly Tyr Ala Cys Arg Pro Gly Ser Ser Ser Ala Asp Gly	380
1141	AGG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAT CCG GAG GCG ATC TAC GAC TCG ATA AGA	1200
381	Arg Pro Val Ser Asp Ile Gly Trp Glu Ile Tyr Pro Glu Gly Ile Tyr Asp Ser Ile Arg	400
1201	GAG GCC AAC AAA TAC GCG CTC CCG GTT TAC GTC ACC GAA AAC GCA ATA GCC CAT TCA ACT	1260
401	Glu Ala Asn Lys Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser Thr	420
1261	GAC ACC CTG CCG CCG TAC TAC CTC GCG ACC CAT GTA CCG AAG ATT GAG GAG CCG TAC GAG	1320
421	Asp Thr Leu Arg Pro Tyr Tyr Leu Ala Ser His Val Ala Lys Ile Glu Glu Ala Tyr Glu	440

Figure 4

1321	CGG GGT TAC GAC GTC ACC GGG TAC CTC TAC TGG GCG CTG ACC GAC AAC TAC GAG TGG GCC	1380
441	Ala Gly Tyr Asp Val Arg Gly Tyr Leu Tyr Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala	460
1381	CTC GGT TTC ACG ATG AGG TTC GGC CTC TAT AAA GTG GAT CTC ATA ACC AAG GAG AGA ACA	1440
461	Leu Gly Phe Arg Met Arg Phe Gly Leu Tyr Lys Val Asp Leu Ile Thr Lys Glu Arg Thr	480
1441	CCG CCG GAG GAA AGC GTA AAG GTT TAT ACC CCG ATC CTC GAG AAC AAC GGA GTC AGC AAC	1500
481	Pro Arg Glu Glu Ser Val Lys Val Tyr Arg Gly Ile Val Glu Asn Asn Gly Val Ser Lys	500
1501	GAA ATC CCG GAG AAG TTC GGA CTT GGG TGA	1530
501	Glu Ile Arg Glu Lys Phe Gly Leu Gly End	520

Figure 4 (Continued)

7/33

1	ATG	GAA	AGG	ATC	GAT	GAA	ATT	CTC	TCT	TAG	TTA	ACT	ACA	GAG	GAA	AAG	GTG	AAG	CTC	GTT	80
1	Met	Glu	Arg	Ile	Asp	Glu	Ile	Leu	Ser	Gln	Leu	Thr	Thr	Glu	Glu	Lys	Val	Lys	Leu	Val	20
61	GTG	GGG	GTT	GGT	CTT	CCA	GGA	CTT	TTT	GUG	AAC	CCA	CAT	TCC	AGA	GTG	GCG	GCT	GCG	GCT	120
21	Val	Gly	Val	Gly	Ile	Pro	Gly	Leu	Phe	Gly	Asn	Pro	His	Ser	Arg	Val	Ala	Gly	Ala	Ala	40
121	GGA	GAA	ACA	CAT	CCC	GTT	CCA	AGA	CTT	GGA	ATT	CCT	GCG	TTT	GTC	CTG	GCA	GAT	GGT	CCC	180
41	Gly	Glu	Thr	His	Pro	Val	Pro	Arg	Leu	Gly	Ile	Pro	Ala	Phe	Val	Leu	Ala	Asp	Gly	Pro	60
181	GCA	GGA	CTC	AGA	ATA	AAT	CCC	ACA	AGG	GAA	AAC	GAT	GAA	AAC	ACT	TAC	TAC	ACG	ACG	GCA	240
61	Ala	Gly	Leu	Arg	Ile	Asn	Pro	Thr	Arg	Glu	Asn	Asp	Glu	Asn	Thr	Tyr	Tyr	Thr	Thr	Ala	80
241	TTT	CCC	GTT	GAA	ATC	ATG	CTC	GCT	TCT	ACC	TGG	AAC	AGA	GAC	CTT	CTG	GAA	GAA	GTG	GGA	300
81	Phe	Pro	Val	Glu	Ile	Met	Leu	Ala	Ser	Thr	Trp	Asn	Arg	Asp	Leu	Leu	Glu	Glu	Val	Gly	100
301	AAA	CCC	ATG	GGA	GAA	GAA	GTT	AGG	GAA	TAC	GGT	GTC	GAT	GTG	CTT	CTT	GCA	CCT	GCG	ATG	360
101	Lys	Ala	Met	Gly	Glu	Glu	Val	Arg	Glu	Tyr	Gly	Val	Asp	Val	Leu	Leu	Ala	Pro	Ala	Met	120
361	AAC	ATT	CAC	AGA	AAC	CCT	CTT	TGT	GGA	AGG	AAT	TTC	GAG	TAC	TAC	TCA	GAA	GAT	CCT	GTC	420
121	Asn	Ile	His	Arg	Asn	Pro	Glu	Cys	Gly	Arg	Asn	Phe	Glu	Tyr	Tyr	Ser	Glu	Asp	Pro	Val	140
421	CTT	TCC	GGT	GAA	ATG	GCT	TCA	GCC	TTT	GTC	AAG	GGA	GTT	CAA	TCT	CAA	GGG	GTG	GGA	GCC	480
141	Leu	Ser	Gly	Glu	Met	Ala	Ser	Ala	Phe	Val	Lys	Gly	Val	Gln	Ser	Gln	Gly	Val	Gly	Ala	160
481	TGC	ATA	AAA	CAC	TTT	GTC	GCG	AAC	AAC	CAG	GAA	ACG	AAC	AGG	ATG	GTA	GTG	GAC	ACG	ATC	540
161	Cys	Ile	Lys	His	Phe	Val	Ala	Asn	Asn	Gln	Glu	Thr	Asn	Arg	Met	Val	Val	Asp	Thr	Ile	180
541	GTG	TCC	GAG	CGA	GCC	CTC	AGA	GAA	ATA	TAT	CTG	AAA	GGT	TTT	GAA	ATT	GCT	GTC	AAG	AAA	600
181	Val	Ser	Glu	Arg	Ala	Leu	Arg	Glu	Ile	Tyr	Leu	Lys	Gly	Phe	Glu	Ile	Ala	Val	Lys	Lys	200
601	GCA	AGA	CCC	TGG	ACC	GTG	ATG	AGC	GCT	TAC	AAC	AAA	CTG	AAT	GGA	AAA	TAC	TGT	TCA	CAG	660
201	Ala	Arg	Pro	Trp	Thr	Val	Met	Ser	Ala	Tyr	Asn	Lys	Leu	Asn	Gly	Lys	Tyr	Cys	Ser	Gln	220
661	AAC	GAA	TGG	CTT	TTG	AAG	AAG	GTT	CTC	AGG	GAA	GAA	TGG	GGA	TTT	GGC	GGT	TTC	GTG	ATG	720
221	Asn	Glu	Trp	Leu	Leu	Lys	Lys	Val	Leu	Arg	Glu	Glu	Trp	Gly	Phe	Gly	Gly	Phe	Val	Met	240
721	AGC	GAC	TGG	TAC	GCG	GGA	GAC	AAC	CCT	GTA	GAA	CAG	CTC	AAG	GCC	GGA	AAC	GAT	ATG	ATC	780
241	Ser	Asp	Trp	Tyr	Ala	Gly	Asp	Asn	Pro	Val	Glu	Gln	Leu	Lys	Ala	Gly	Asn	Asp	Met	Ile	260
781	ATG	CCT	GGG	AAA	GCG	TAT	CAG	GTG	AAC	ACA	GAA	AGA	AGA	GAT	GAA	ATA	GAA	GAA	ATC	ATG	840
261	Met	Pro	Gly	Lys	Ala	Tyr	Gln	Val	Asn	Thr	Glu	Arg	Arg	Asp	Glu	Ile	Glu	Glu	Ile	Met	280
841	GAG	GCG	TTG	AAG	GAG	GGA	AAA	TTG	AGT	GAG	GAG	GTT	CTC	GAT	GAG	TGT	GTG	AGA	AAC	ATT	900
281	Gln	Ala	Leu	Lys	Glu	Gly	Lys	Leu	Ser	Glu	Glu	Val	Leu	Asp	Glu	Cys	Val	Arg	Asn	Ile	300
901	CTC	AAA	GTT	CTT	GTG	AAC	GCG	CCT	TCC	TTC	AAA	GGG	TAC	AGG	TAC	TCA	AAC	AAG	CCG	GAT	960
301	Leu	Lys	Val	Leu	Val	Asn	Ala	Pro	Ser	Phe	Lys	Gly	Tyr	Arg	Tyr	Ser	Asn	Lys	Pro	Asp	320
961	CTC	GAA	TCT	CAC	GCG	GAA	GTC	GCC	TAC	GAA	GCA	GGT	GCG	GAG	GGT	GTT	GTC	CTT	CTT	GAG	1020
321	Leu	Glu	Ser	His	Ala	Glu	Val	Ala	Tyr	Glu	Ala	Gly	Ala	Glu	Gly	Val	Val	Leu	Leu	Glu	340
1021	AAC	AAC	GGT	GTT	CTT	CCG	TTC	GAT	GAA	AAT	ACC	TAT	GTC	GCC	GTC	TTT	GGC	ACC	GGT	CAA	1080
341	Asn	Asn	Gly	Val	Leu	Pro	Phe	Asp	Glu	Asn	Thr	His	Val	Ala	Val	Phe	Gly	Thr	Gly	Gln	360
1081	ATC	GAA	ACA	ATA	AAG	GGA	GGA	ACG	GGA	AGT	GGA	GAC	ACC	CAT	CCG	AGA	TAC	ACG	ATC	TCT	1140
361	Ile	Glu	Thr	Ile	Lys	Gly	Gly	Thr	Gly	Ser	Gly	Asp	Thr	His	Pro	Arg	Tyr	Thr	Ile	Ser	380
1141	ATC	CTT	GAA	GAG	ATA	AAA	GAA	AGA	AAC	ATG	AAG	TTT	GAC	GAA	GAA	CTC	GCT	TCC	ACT	TAT	1200
381	Ile	Leu	Glu	Gly	Ile	Lys	Glu	Arg	Asn	Met	Lys	Phe	Asp	Glu	Glu	Leu	Ala	Ser	Thr	Tyr	400

Figure 5

8/33

1201 401	GAG Glu	GAG Glu	TAC Tyr	ATA Ile	AAA Lys	AAG Lys	ATG Met	ATA Ile	GAA Glu	ACA Thr	GAG Glu	GAA Glu	TAT Tyr	AAA Lys	CTT Pro	AGA Arg	ACT Thr	GAT Asp	TCT Ser	TGG Trp	1281 420
1261 421	GGA Gly	ACG Thr	GTC Val	ATA Ile	AAA Lys	CCG Pro	AAA Lys	CTC Leu	CCA Pro	GAG Glu	AAT Asn	TTC Phe	CTC Leu	TCA Ser	GAA Glu	AAA Lys	GAG Glu	ATA Ile	AAG Lys	AAA Lys	1320 440
1321 441	CCT Pro	CCA Pro	AAG Lys	AAA Lys	AAC Asn	GAT Asp	GTT Val	GCA Ala	GTT Val	GTT Val	GTG Val	ATC Ile	AGT Ser	AGG Arg	ATC Ile	TCC Ser	GGT Gly	GAG Glu	GGA Gly	TAC Tyr	1380 460
1381 461	GAC Asp	AGA Arg	AAG Lys	CCG Pro	GTC Val	AAA Lys	GGT Gly	GAC Asp	TTC Phe	TAC Tyr	CTC Leu	TCC Ser	GAT Asp	GAC Asp	GAG Glu	CTG Leu	GAA Glu	CTC Leu	ATA Ile	AAA Lys	1440 480
1441 481	ACC Thr	GTC Val	TCG Ser	AAA Lys	GAA Glu	TTC Phe	CAC His	GAT Asp	CAG Gln	GGT Gly	AAG Lys	AAA Lys	GTT Val	GTG Val	GTT Val	CTT Leu	CTG Leu	AAC Asn	ATC Ile	GGA Gly	1500 500
1501 501	AGT Ser	CCC Pro	ATC Ile	GAA Glu	GTC Val	GCA Ala	AGC Ser	TGG Trp	AGA Arg	GAC Asp	CTT Leu	GTG Val	GAT Asp	GGA Gly	ATT Ile	CTT Leu	CTC Leu	GTC Val	TGG Trp	CAG Gln	1560 520
1561 521	GCG Ala	GGA Gly	CAG Gln	GAG Met	ATG Met	GGA Gly	AGA Arg	ATA Ile	GTG Val	GCC Ala	GAT Asp	GTT Val	CTT Leu	GTG Val	GGA Gly	AAG Lys	ATT Ile	AAT Asn	CCC Pro	TCC Ser	1620 540
1621 541	GGA Gly	AAA Lys	CTT Leu	CCA Pro	ACG Thr	ACC Thr	TTC Phe	CCG Pro	AAG Lys	GAT Asp	TAC Tyr	TCG Ser	GAC Asp	GTT Val	CCA Pro	TCC Ser	TGG Trp	ACG Thr	TTC Phe	CCA Pro	1680 560
1681 561	GGA Gly	GAG Glu	CCA Pro	AAG Lys	GAC Asp	AAT Asn	CCG Pro	CAA Gln	AGA Arg	GTG Val	GTG Val	TAC Tyr	GAG Glu	GAA Glu	GAC Asp	ATC Ile	TAC Tyr	GTG Val	GGA Gly	TAC Tyr	1740 580
1741 581	AGG Arg	TAC Tyr	TAC Tyr	GAC Asp	ACC Thr	TTC Phe	GGT Gly	GTG Val	GAA Glu	CCT Pro	GCC Ala	TAC Tyr	GAA Glu	TTC Phe	GGC Gly	TAC Tyr	GGC Gly	CTC Leu	TCT Ser	TAC Tyr	1800 600
1801 601	ACA Thr	AAG Lys	TTT Phe	GAA Glu	TAC Tyr	AAA Lys	GAT Asp	TTA Leu	AAA Lys	ATC Ile	GCT Ala	ATC Ile	GAC Asp	GGT Gly	GAG Glu	ACG Thr	CTC Leu	AGA Arg	GTG Val	TCG Ser	1860 620
1861 621	TAC Tyr	ACG Thr	ATC Ile	ACA Thr	AAC Asn	ACT Thr	GGG Gly	GAC Asp	AGA Arg	GCT Ala	GGA Gly	AAG Lys	GAA Glu	GTC Val	TCA Ser	CAG Gln	GTC Val	TAC Tyr	ATC Ile	AAA Lys	1920 640
1921 641	GCT Ala	CCA Pro	AAA Lys	GGA Gly	AAA Lys	ATA Ile	GAC Asp	AAA Lys	CCC Pro	TTC Phe	CAG Gln	GAG Glu	CTG Leu	AAA Lys	GCG Ala	TTT Phe	CAC His	AAA Lys	ACA Thr	AAA Lys	1980 660
1981 661	CTT Leu	TTG Leu	AAC Asn	CCG Pro	GGT Gly	GAA Glu	TCA Ser	GAA Glu	ATC Ile	TCC Ser	TTG Leu	GAA Glu	ATT Ile	CCT Pro	CTC Leu	AGA Arg	GAT Asp	CTT Leu	GCG Ala	2040 680	
2041 681	AGT Ser	TTC Phe	GAT Asp	GGG Gly	AAA Lys	GAA Glu	TGG Trp	GTT Val	GTC Val	GAG Glu	TCA Ser	GGA Gly	GAA Glu	TAC Tyr	GAG Glu	GTC Val	AGG Arg	GTC Val	GGT Gly	GCA Ala	2100 700
2101 701	TCT Ser	TCG Ser	AGG Arg	GAT Asp	ATA Ile	AGG Arg	TTG Leu	AGA Arg	GAT Asp	ATT Ile	TTT Phe	CTG Leu	GTT Val	GAG Glu	GGA Gly	GAG Glu	AAG Lys	AGA Arg	TTC Phe	AAA Lys	2160 720
2161 721	CCA Pro	TGA End																			2166 722

Figure 5 (Continued)



## THERMOCOCCUS AEDII12RA GLYCOSIDASE (18B/G)

## COMPLETE GENE SEQUENCE - 9/95

1	ATG ATC CAC TGC CCG GTT AAA GCG ATT ATA TCT GAG GCT CCC GGC ATA ACC ATC ACA ATA	60
1	Met Ile His Cys Pro Val Lys Gly Ile Ile Ser Glu Ala Arg Gly Ile Thr Ile Thr Ile	20
61	GAT TTA AGT TTT CAA GGC CAA ATA AAT AAT TTG GTG AAT GCT ATG ATT GTC TTT CCG GAG	120
21	Asp Leu Ser Phe Gln Gly Gln Ile Asn Asn Leu Val Asn Ala Met Ile Val Phe Pro Glu	40
121	TTC TTC CTC TTT GGA ACC GCC ACA TCT TCT CAT CAG ATC GAG GGA GAT AAT AAA TGG AAC	180
41	Phe Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile Glu Gly Asp Asn Lys Trp Asn	60
181	GAC TGG TGG TAT TAT GAG GAG ATA GGT AAG CTC CCC TAC AAA TCC GGT AAA GCC TGC AAT	240
61	Asp Trp Trp Tyr Tyr Glu Glu Ile Gly Lys Leu Pro Tyr Lys Ser Gly Lys Ala Cys Asn	80
241	CAC TGG GAG CTT TAC AGG GAA GAT ATA GAG CTA ATG GCA CAG CTC GCC TAC AAT GCC TAC	300
81	His Trp Glu Leu Tyr Arg Glu Asp Ile Glu Leu Met Ala Gln Leu Gly Tyr Asn Ala Tyr	100
301	CGC TTT TCG ATA GAG TGG AGC CGT CTC TTC CCG GAA GAG GCC AAA TTC AAT GAA GAA GCC	360
101	Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu Gly Lys Phe Asn Glu Glu Ala	120
361	TTC AAC CGC TAC CGT GAA ATA ATT GAA ATC CTC CTT GAG AAG GGG ATT ACT CCA AAC GTT	420
121	Phe Asn Arg Tyr Arg Glu Ile Ile Glu Ile Leu Leu Glu Lys Gly Ile Thr Pro Asn Val	140
421	ACA CTG CAC CAC TTC ACA TCA CCG CTG TGG TTC ATG CCG AAG GGA GCC TTT TTG AAG GAA	480
141	Thr Leu His His Phe Thr Ser Pro Leu Trp Phe Met Arg Lys Gly Gly Phe Leu Lys Glu	160
481	GAA AAC CTC AAG TAC TGG GAG CAG TAC GTT GAT AAA GCC GCG GAG CTC CTC AAG GGA GTC	540
161	Glu Asn Leu Lys Tyr Trp Glu Gln Tyr Val Asp Lys Ala Ala Glu Leu Leu Lys Gly Val	180
541	AAG CTT GTA GCT ACA TTC AAC GAG CCG ATG GTC TAT GTT ATG ATG GCC TAC CTC ACA GCC	600
181	Lys Leu Val Ala Thr Phe Asn Glu Pro Met Val Tyr Val Met Met Gly Tyr Leu Thr Ala	200
601	TAC TGG CCG CCC TTC ATC AAG AGT CCC TTT AAA GCC TTT AAA GTT GCC CCA AAC CTC CTT	660
201	Tyr Trp Pro Pro Phe Ile Lys Ser Pro Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu	220
661	AAG GCC CAT GCA ATG GCA TAT GAT ATC CTC CAT GGT AAC TTT GAT GTG GGG ATA GTT AAA	720
221	Lys Ala His Ala Met Ala Tyr Asp Ile Leu His Gly Asn Phe Asp Val Gly Ile Val Lys	240
721	AAC ATC CCC ATA ATG CTC CCT GCA AGC AAC AGA GAG AAA GAC GTA GAA GCT GCC CAA AAG	780
241	Asn Ile Pro Ile Met Leu Pro Ala Ser Asn Arg Glu Lys Asp Val Glu Ala Ala Gln Lys	260
781	CGC GAT AAC CTC TTT AAC TGG AAC TTC CTT GAT GCA ATA TGG AGC GGA AAA TAT AAA GGA	840
261	Ala Asp Asn Leu Phe Asn Trp Asn Phe Leu Asp Ala Ile Trp Ser Gly Lys Tyr Lys Gly	280
841	GCT TTT GGA ACT TAC AAA ACT CCA GAA AGC GAT GCA GAC TTC ATA GGG ATA AAC TAC TAC	900
281	Ala Phe Gly Thr Tyr Lys Thr Pro Glu Ser Asp Ala Asp Phe Ile Gly Ile Asn Tyr Tyr	300
901	ACA GCC AGC GAG GTA AGG CAT AGC TGG AAT CCG CTA AAG TTT TTC TTC GAT GCC AAG CTT	960
301	Thr Ala Ser Glu Val Arg His Ser Trp Asn Pro Leu Lys Phe Phe Phe Asp Ala Lys Leu	320
961	GCA GAC TTA AGC GAG AGA AAA ACA GAT ATG GGT TGG AGT GTC TAT CCA AAG GCC ATA TAC	1020
321	Ala Asp Leu Ser Glu Arg Lys Thr Asp Met Gly Trp Ser Val Tyr Pro Lys Gly Ile Tyr	340
1021	GAA GCT ATA GCA AAG GTT TCA CAC TAC GGA AAG CCA ATG TAC ATC ACG GAA AAC GGG ATA	1080
341	Glu Ala Ile Ala Lys Val Ser His Tyr Gly Lys Pro Met Tyr Ile Thr Glu Asn Gly Ile	360
1081	GCT ACC TTA GAC GAT GAG TGG AGG ATA GAG TTT ATC ATC CAG CAC CTC CAG TAC GTT CAC	1140
361	Ala Thr Leu Asp Asp Glu Trp Arg Ile Glu Phe Ile Ile Gln His Leu Gln Tyr Val His	380
1141	AAA GCC TTA AAC GAT GGC TTT GAC TTG AGA GGC TAC TTC TAT TGG TCT TTT ATG GAT AAC	1200
381	Lys Ala Leu Asn Asp Gly Phe Asp Leu Arg Gly Tyr Phe Tyr Trp Ser Phe Met Asp Asn	400
1201	TTC GAG TGG GCT GAG GGT TTT AGA CCA CCG TTT GGG CTG GTC GAG GTG GAC TAC ACG ACC	1260
401	Phe Glu Trp Ala Glu Gly Phe Arg Pro Arg Phe Gly Leu Val Glu Val Asp Tyr Thr Thr	420
1261	TTC AAG AGG AGA CCG AGA AAG AGT GCT TAC ATA TAT GGA GAA ATT GCA AGG GAA AAG AAA	1320
421	Phe Lys Arg Arg Pro Arg Lys Ser Ala Tyr Ile Tyr Gly Glu Ile Ala Arg Glu Lys Lys	440
1321	ATA AAA GAC GAA CTG CTG GCA AAG TAT GGG CTT CCG GAG CTA TGA	1365
441	Ile Lys Asp Glu Leu Leu Ala Lys Tyr Gly Leu Pro Glu Leu End	455

Figure 6

THERMOCOCCUS CHITONOPHAGUS GLYCOSIDASE - 220  
COMPLETE SEQUENCE - 9/95

1	TTG CTT CCA GAG AAC TTT CTC TGG GGA GTT TCA CAG TCC GGA TTC CAG TTT GAA ATG GCG	60
1	Met Leu Pro Glu Asn Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Gln Phe Glu Met Gly	20
61	GAC AGA CTG AGG AGG CAC ATT GAT CCA AAC ACA GAT TGG TGG TAC TGG GTA AGA GAT GAA	120
21	Asp Arg Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp Tyr Trp Val Arg Asp Glu	40
121	TAT AAT ATC AAA AAA GGA CTA GTA AGT GGG GAT CTT CCC GAA GAC GGT ATA AAT TCA TAT	180
41	Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Glu Asp Gly Ile Asn Ser Tyr	60
181	GAA TTA TAT GAG AGA GAC CAA GAA ATT GCA AAG GAT TTA GGG CTC AAC ACA TAT AGG ATC	240
61	Glu Leu Tyr Glu Arg Asp Gln Glu Ile Ala Lys Asp Leu Gly Leu Asn Thr Tyr Arg Ile	80
241	GGA ATT GAA TGG AGC AGA GTA TTT CCA TGG CCA ACG ACT TTT GTC GAC GTG GAG TAT GAA	300
81	Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Asp Val Glu Tyr Glu	100
301	ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG AAG ATT TCT AAA GAC GCA TTA GAA AAA	360
101	Ile Asp Glu Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Ser Lys Asp Ala Leu Glu Lys	120
361	CTT GAT GAA ATC GCT AAC CAA AGG GAA ATA ATA TAT TAT AGG AAC CTA ATA AAT TCC CTA	420
121	Leu Asp Glu Ile Ala Asn Gln Arg Glu Ile Ile Tyr Tyr Arg Asn Leu Ile Asn Ser Leu	140
421	AGA AAG AGG GGT TTT AAG GTA ATA CTA AAC CTA AAT CAT TTT ACC CTC CCA ATA TGG CTT	480
141	Arg Lys Arg Gly Phe Lys Val Ile Leu Asn Leu Asn His Phe Thr Leu Pro Ile Trp Leu	160
481	CAT GAT CCT ATC GAA TCT AGA GAA AAA GCC CTG ACC AAT AAG AGA AAC GGA TGG GTA AGC	540
161	His Asp Pro Ile Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg Asn Gly Trp Val Ser	180
541	GAA AGG AGT GTT ATA GAG TTT GCA AAA TTT GCC GCG TAT TTA GCA TAT AAA TTC GGA GAC	600
181	Glu Arg Ser Val Ile Glu Phe Ala Lys Phe Ala Ala Tyr Leu Ala Tyr Lys Phe Gly Asp	200
601	ATA GTA GAC ATG TGG AGC ACA TTT AAT GAA CCT ATG GTG GTC GCC GAG TTG GGG TAT TTA	660
201	Ile Val Asp Met Trp Ser Thr Phe Asn Glu Pro Met Val Val Ala Glu Leu Gly Tyr Leu	220
661	GCC CCA TAC TCA GGA TTC CCC CCG GGA GTC ATG AAT CCA GAA GCA GCA AAG TTA GTT ATG	720
221	Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Met Asn Pro Glu Ala Ala Lys Leu Val Met	240
721	CTA CAT ATG ATA AAC GCC CAT GCT TTA GCA TAT AGG ATG ATA AAG AAA TTT GAC AGA AAA	780
241	Leu His Met Ile Asn Ala His Ala Leu Ala Tyr Arg Met Ile Lys Lys Phe Asp Arg Lys	260
781	AAA GCT GAT CCA GAA TCA AAA GAA CCA GCT GAA ATA GGA ATT ATA TAC AAT AAC ATC GGC	840
261	Lys Ala Asp Pro Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile Ile Tyr Asn Asn Ile Gly	280
841	GTC ACA TAT CCG TTT AAT CCG AAA GAC TCA AAG GAT CTA CAA GCA TCC GAT AAT GCC AAT	900
281	Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gln Ala Ser Asp Asn Ala Asn	300
901	TTC TTC CAC AGT GGG CTA TTC TTA ACG GCT ATC CAC AGG GGA AAA TTA AAT ATC GAA TTT	960
301	Phe Phe His Ser Gly Leu Phe Leu Thr Ala Ile His Arg Gly Lys Leu Asn Ile Glu Phe	320
961	GAC GGA GAG ACA TTT GTT TAC CTT CCA TAT TTA AAG GGC AAT GAT TGG CTG GGA GTG AAT	1020
321	Asp Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn	340
1021	TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TTT CCA AGT ATC CCT CTC ATA	1080
341	Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Met Phe Pro Ser Ile Pro Leu Ile	360
1081	AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC	1140
361	Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp	380
1141	GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA	1200
381	Gly Asn Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Lys Gly Met Tyr Asp Ser Ile	400
1201	GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA	1260
401	Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser	420
1261	AAA GAT GTA TTA AGG CCC TAT TAC ATC GCA TCT CAC ATT GAA GCC ATG GAA GAG GCT TAC	1320
421	Lys Asp Val Leu Arg Pro Tyr Tyr Ile Ala Ser His Ile Glu Ala Met Glu Glu Ala Tyr	440

Figure 7

1321	GAA AAT GGT TAT GAC GTG AGA GGA TAC TTA CAC TGG GCA TTA ACC GAT AAT TAC GAA TGG	1380
441	Glu Asn Gly Tyr Asp Val Arg Gly Tyr Leu His Trp Ala Leu Thr Asp Asn Tyr Glu Trp	460
1381	GCC TTA GGG TTC AGA ATG AGG TTT GGC TTG TAC GAA GTA AAC TTG ATA ACC AAA GAG AGA	1440
461	Ala Leu Gly Phe Arg Met Arg Phe Gly Leu Tyr Glu Val Asn Leu Ile Thr Lys Glu Arg	480
1441	AAA CCC AGG AAA AAG AGT GTA AGA GTA TTC AGA GAG ATA GTT ATT AAT AAT GGG CTA ACA	1500
481	Lys Pro Arg Lys Lys Ser Val Arg Val Phe Arg Glu Ile Val Ile Asn Asn Gly Leu Thr	500
1501	AGC AAC ATC AGG AAA GAG ATC TTA GAG GAG GGG TAG	1536
501	Ser Asn Ile Arg Lys Glu Ile Leu Glu Glu Gly End	512

Figure 7 (Continued)

12/33

**PYROCOCCUS FURIOSUS GLYCOSIDASE - 7G1**  
**COMPLETE GENE SEQUENCE - 10/95**

1	ATG TTC CCT GAA AAG TTC CTT TGG GGT GTG GCA CAA TCG GGT TTT CAG TTT GAA ATG GGG	60
1	Met Phe Pro Glu Lys Phe Leu Trp Gly Val Ala Gln Ser Gly Phe Gln Phe Glu Met Gly	20
61	GAT AAA CTC AGG AGG AAT ATT GAC ACT AAC ACT GAT TGG TGG CAC TCG GTA AGG GAT AAG	120
22	Asp Lys Leu Arg Arg Asn Ile Asp Thr Asn Thr Asp Trp Trp His Trp Val Arg Asp Lys	40
121	ACA AAT ATA GAG AAA GGC CTC GTT AGT GGA GAT CTT CCC GAG GAG GGG ATT AAC AAT TAC	180
41	Thr Asn Ile Glu Lys Gly Leu Val Ser Gly Asp Leu Pro Glu Glu Gly Ile Asn Asn Tyr	60
181	GAG CTT TAT GAG AAG GAC CAT GAG ATT CCA AGA AAG CTG GGT CTT AAT GCT TAC AGA ATA	240
61	Glu Leu Tyr Glu Lys Asp His Glu Ile Ala Arg Lys Leu Gly Leu Asn Ala Tyr Arg Ile	80
241	GGC ATA GAG TGG AGC AGA ATA TTC CCA TGG CCA ACG ACA TTT ATT GAT GTT GAT TAT AGC	300
81	Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp Pro Thr Thr Phe Ile Asp Val Asp Tyr Ser	100
301	TAT AAT GAA TCA TAT AAC CTT ATA GAA GAT GTA AAG ATC ACC AAG GAC ACT TTG GAG GAG	360
101	Tyr Asn Glu Ser Tyr Asn Leu Ile Glu Asp Val Lys Ile Thr Lys Asp Thr Leu Glu Glu	120
361	TTA GAT GAG ATC GCC AAC AAG AGG GAG GTC GCC TAC TAT AGG TCA GTC ATA AAC AGC CTG	420
121	Leu Asp Glu Ile Ala Asn Lys Arg Glu Val Ala Tyr Tyr Arg Ser Val Ile Asn Ser Leu	140
421	AGG AGC AAG GGG TTT AAG GTT ATA GTT AAT CTA AAT CAC TTC ACC CTT CCA TAT TGG TTG	480
141	Arg Ser Lys Gly Phe Lys Val Ile Val Asn Leu Asn His Phe Thr Leu Pro Tyr Trp Leu	160
481	CAT GAT CCC ATT GAG GCT AGG GAG AGG GCG TTA ACT AAT AAG AGG AAC GGC TGG GTT AAC	540
161	His Asp Pro Ile Glu Ala Arg Glu Arg Ala Leu Thr Asn Lys Arg Asn Gly Trp Val Asn	180
541	CCA AGA ACA GTT ATA GAG TTT GCA AAG TAT GCC GCT TAC ATA GCC TAT AAG TTT GGA GAT	600
181	Pro Arg Thr Val Ile Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala Tyr Lys Phe Gly Asp	200
601	ATA GTG GAT ATG TGG AGC ACG TTT AAT GAG CCT ATG GTG GTT GTT GAG CTT GGC TAC CTA	660
201	Ile Val Asp Met Trp Ser Thr Phe Asn Glu Pro Met Val Val Val Glu Leu Gly Tyr Leu	220
661	GCC CCC TAC TCT GGC TTC CCT CCA GGG GTT CTA AAT CCA GAG GCC GCA AAG CTG GCG ATA	720
221	Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Leu Asn Pro Glu Ala Ala Lys Leu Ala Ile	240
721	CTT CAC ATG ATA AAT GCA CAT GCT TTA GCT TAT AGG CAG ATA AAG AAG TTT GAC ACT GAG	780
241	Leu His Met Ile Asn Ala His Ala Leu Ala Tyr Arg Gln Ile Lys Lys Phe Asp Thr Glu	260
781	AAA GCT GAT AAG GAT TCT AAA GAG CCT GCA GAA GTT GGT ATA ATT TAC AAC AAC ATT GGA	840
261	Lys Ala Asp Lys Asp Ser Lys Glu Pro Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly	280
841	GTT GCT TAT CCC AAG GAT CCG AAC GAT TCC AAG GAT GTT AAG GCA GCA GAA AAC GAC AAC	900
281	Val Ala Tyr Pro Lys Asp Pro Asn Asp Ser Lys Asp Val Lys Ala Ala Glu Asn Asp Asn	300
901	TTC TTC CAC TCA GGG CTG TTC TTC GAG GCC ATA CAC AAA GGA AAA CTT AAT ATA GAG TTT	960
301	Phe Phe His Ser Gly Leu Phe Phe Glu Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe	320
961	GAC GGT GAA ACG TTT ATA GAT GCC CCC TAT CTA AAG GGC AAT GAC TCG ATA GGG GTT AAT	1020
321	Asp Gly Glu Thr Phe Ile Asp Ala Pro Tyr Leu Lys Gly Asn Asp Trp Ile Gly Val Asn	340
1021	TAC TAC ACA AGG GAA GTA GTT ACG TAT CAG GAA CCA ATG TTT CCT TCA ATC CCG CTG ATC	1080
341	Tyr Tyr Thr Arg Glu Val Val Thr Tyr Gln Glu Pro Met Phe Pro Ser Ile Pro Leu Ile	360
1081	ACC TTT AAG GGA GTT CAA GGA TAT GGC TAT GCC TGC AGA CCT GGA ACT CTG TCA AAG GAT	1140
361	Thr Phe Lys Gly Val Gln Gly Tyr Gly Tyr Ala Cys Arg Pro Gly Thr Leu Ser Lys Asp	380
1141	GAC AGA CCC GTC AGC GAC ATA GGA TGG GAA CTC TAT CCA GAG GGG ATG TAC GAT TCA ATA	1200
381	Asp Arg Pro Val Ser Asp Ile Gly Trp Glu Leu Tyr Pro Glu Gly Met Tyr Asp Ser Ile	400
1201	GTT GAA GCT CAC AAG TAC GGC GTT CCA GTT TAC GTG ACG GAG AAC GGA ATA GCG GAT TCA	1260
401	Val Glu Ala His Lys Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser	420

Figure 8

1261	AAG GAC ATC CTA AGA CCT TAC TAC ATA GCG AGC CAC ATA AAG ATG ATA GAG AAG GCC TTT	1320
421	Lys Asp Ile Leu Arg Pro Tyr Tyr Ile Ala Ser His Ile Lys Met Ile Glu Lys Ala Phe	440
1321	GAG GAT GGG TAT GAA GTT AAG GGC TAC TTC CAC TGG GCA TTA ACT GAC AAC TTC GAG TGG	1380
441	Glu Asp Gly Tyr Glu Val Lys Gly Tyr Phe His Trp Ala Leu Thr Asp Asn Phe Glu Trp	460
1381	GCT CTC GGG TTT AGA ATG CGC TTT GGC CTC TAC GAA GTC AAC CTA ATT ACA AAG GAG AGA	1440
461	Ala Leu Gly Phe Arg Met Arg Phe Gly Leu Tyr Glu Val Asn Leu Ile Thr Lys Glu Arg	480
1441	ATT CCC AGG GAG AAG AGC GTG TCG ATA TTC AGA GAG ATA GTA GCC AAT AAT GGT GTT ACG	1500
481	Ile Pro Arg Glu Lys Ser Val Ser Ile Phe Arg Glu Ile Val Ala Asn Asn Gly Val Thr	500
1501	AAA AAG ATT GAA GAG GAA TTG CTG AGG GGA TGA	1533
501	Lys Lys Ile Glu Glu Glu Leu Leu Arg Gly End	511

Figure 8 (Continued)



**Bankia gouldi endoglucanase (370F1) (continued)**

711	720	729	738	747	756
CTG CAC ACC TAT TTC GAA ACC GCC AAA AAA GCC CGC GCC AAA TTT CCC GGT ATT					
Leu His Thr Tyr Phe Glu Thr Ala Lys Lys Ala Arg Ala Lys Phe Pro Gly Ile					
765	774	783	792	801	810
AAA ATC ACC GGT CCG GTG CCC GCT AAT GAG TGG CAG TGG TAT GCC TGG GCC GGT					
Lys Ile Thr Gly Pro Val Pro Ala Asn Glu Trp Gln Trp Tyr Ala Trp Gly Gly					
819	828	837	846	855	864
TTC TCG GTA CCC CAG GAA CAA GCG TTT ATG AGC TGG ATG GAG TAT TTC ATC AAG					
Phe Ser Val Pro Gln Glu Gln Gly Phe Met Ser Trp Met Glu Tyr Phe Ile Lys					
873	882	891	900	909	918
CGG GTG TCT GAA GAG CAA CCG GCA AGT GGT GTT CCG CTC CTC GAT GTA CTC GAT					
Arg Val Ser Glu Glu Gln Arg Ala Ser Gly Val Arg Leu Leu Asp Val Leu Asp					
927	936	945	954	963	972
CTG CAC TAC TAC CCC GGC GCT TAC AAT GCG GAA GAT ATC GTG CAA TTA CAT CCG					
Leu His Tyr Tyr Pro Gly Ala Tyr Asn Ala Glu Asp Ile Val Gln Leu His Arg					
981	990	999	1008	1017	1026
ACG TTC TTC GAC CCG GAC TTT GTT TCA CTG GAT CCC AAC GGG GTG AAA ATG GTA					
Thr Phe Phe Asp Arg Asp Phe Val Ser Leu Asp Ala Asn Gly Val Lys Met Val					
1035	1044	1053	1062	1071	1080
GAA GGT GGC TGG GAT GAC AGC ATC AAC AAG GAA TAT ATT TTC GGC CGA GTG AAC					
Glu Gly Gly Trp Asp Asp Ser Ile Asn Lys Glu Tyr Ile Phe Gly Arg Val Asn					
1089	1098	1107	1116	1125	1134
GAT TGG CTC GAG GAA TAT ATG GCG CCA GAC CAT GGT GTA ACC CTG GGC TTA ACC					
Asp Trp Leu Glu Glu Tyr Met Gly Pro Asp His Gly Val Thr Leu Gly Leu Thr					
1143	1152	1161	1170	1179	1188
GAA ATG TGC GTG CCG AAT GTG AAT CCG ATG ACT ACC GCC ATC TGG TAT GCC TCC					
Glu Met Cys Val Arg Asn Val Asn Pro Met Thr Thr Ala Ile Trp Tyr Ala Ser					
1197	1206	1215	1224	1233	1242
ATG CTC GGC ACC TTC GCG GAT AAC GCG GTC GAA ATA TTC ACC CCA TGG TGC TGG					
Met Leu Gly Thr Phe Ala Asp Asn Gly Val Glu Ile Phe Thr Pro Trp Cys Trp					
1251	1260	1269	1278	1287	1296
AAC ACC GGA ATG TGG GAA ACA CTC CAC CTC TTC AGC CCG TAC AAC AAA CCT TAT					
Asn Thr Gly Met Trp Glu Thr Leu His Leu Phe Ser Arg Tyr Asn Lys Pro Tyr					
1305	1314	1323	1332	1341	1350
CGG GTC GCC TCC AGC TCC AGT CTT GAA GAG TTT GTC AGC GCC TAC AGC TCC ATT					
Arg Val Ala Ser Ser Ser Ser Leu Glu Glu Phe Val Ser Ala Tyr Ser Ser Ile					
1359	1368	1377	1386	1395	1404
AAC GAA GCA GAA GAC GCC ATG ACG GTA CTT CTG GTG AAT CGT TCC ACT AGC GAG					
Asn Glu Ala Glu Asp Ala Met Thr Val Leu Leu Val Asn Arg Ser Thr Ser Glu					

Figure 9 (Continued)

**Bankia gonidi endoglucanase (37GP1) (continued)**

1413	1422	1431	1440	1449	1458
ACC CAC ACC GCC ACT GTC GCT ATC GAC GAT TTC CCA CTG GAT GGC CCC TAC CGC					
Thr His Thr Ala Thr Val Ala Ile Asp Asp Phe Pro Leu Asp Gly Pro Tyr Arg					
1467	1476	1485	1494	1503	1512
ACC CTG CGC TTA CAC AAC CTG CCG GGG GAG GAA ACC TTC GTA TCT CAC CGA GAC					
Thr Leu Arg Leu His Asn Leu Pro Gly Glu Glu Thr Phe Val Ser His Arg Asp					
1521	1530	1539	1548	1557	1566
AAC GCC CTG GAA AAA GGT ACA GTG CGC GCC AGC GAC AAT ACG GTA ACA CTG GAG					
Asn Ala Leu Glu Lys Gly Thr Val Arg Ala Ser Asp Asn Thr Val Thr Leu Glu					
1575	1584	1593	1602	1611	
TTG CCC CCT CTG TCC GTT ACT GCA ATA TTG CTC AAG GCC CGG CCC TAA 3'					
Leu Pro Pro Leu Ser Val Thr Ala Ile Leu Leu Lys Ala Arg Pro ***					

Figure 9 (Continued)



17/33

*Thermotoga maritima* Alpha-galactosidase  
Complete Gene Sequence (1043)

```

5'  GTG ATC TGT GTG GAA ATA TTC GGA AAG ACC TTC AGA GAG GGA AGA TTC GTT CTC
    Val Ile Cys Val Glu Ile Phe Gly Lys Thr Phe Arg Glu Gly Arg Phe Val Leu

      9      18      27      36      45      54
    AAA GAG AAA AAC TTC ACA GTT GAG TTC GCG GTG GAG AAG ATA CAC CTT GCG TCG
    Lys Glu Lys Asn Phe Thr Val Glu Phe Ala Val Glu Lys Ile His Leu Gly Trp

      63      72      81      90      99     108
    AAG ATC TCC GGC AGG GTG AAG GGA AGT CCG GGA AGG CTT GAG GTT CTT CGA ACG
    Lys Ile Ser Gly Arg Val Lys Gly Ser Pro Gly Arg Leu Glu Val Leu Arg Thr

     117     126     135     144     153     162
    AAA GCA CCG GAA AAG GTA CTT GTG AAC AAC TGG CAG TCC TGG GGA CCG TGC AGG
    Lys Ala Pro Glu Lys Val Leu Val Asn Asn Trp Gln Ser Trp Gly Pro Cys Arg

     171     180     189     198     207     216
    GTG GTC GAT GCC TTT TCT TTC AAA CCA CCT GAA ATA GAT CCG AAC TGG AGA TAC
    Val Val Asp Ala Phe Ser Phe Lys Pro Pro Glu Ile Asp Pro Asn Trp Arg Tyr

     225     234     243     252     261     270
    ACC GCT TCG GTG GTG CCC GAT GTA CTT GAA AGG AAC CTC CAG AGC GAC TAT TTC
    Thr Ala Ser Val Val Pro Asp Val Leu Glu Arg Asn Leu Gln Ser Asp Tyr Phe

     279     288     297     306     315     324
    GTG OCT GAA GAA GGA AAA GTG TAC GGT TTT CTG AGT TCG AAA ATC GCA CAT CCT
    Val Ala Glu Glu Gly Lys Val Tyr Gly Phe Leu Ser Ser Lys Ile Ala His Pro

     333     342     351     360     369     378
    TTC TTC GCT GTG GAA GAT GCG GAA CTT GTG GCA TAC CTC GAA TAT TTC GAT GTC
    Phe Phe Ala Val Glu Asp Gly Glu Leu Val Ala Tyr Leu Glu Tyr Phe Asp Val

     387     396     405     414     423     432
    GAG TTC GAC GAC TTT GTT CCT CTT GAA CCT CTC GTT GTA CTC GAG GAT CCC AAC
    Glu Phe Asp Asp Phe Val Pro Leu Glu Pro Leu Val Val Leu Glu Asp Pro Asn

     441     450     459     468     477     486
    ACA CCC CTT CTT CTG GAG AAA TAC GCG GAA CTC GTC GGA ATG GAA AAC AAC GCG
    Thr Pro Leu Leu Leu Glu Lys Tyr Ala Glu Leu Val Gly Met Glu Asn Asn Ala

     495     504     513     522     531     540
    AGA GTT CCA AAA CAC ACA CCC ACT GGA TCG TGC ACC TCG TAC CAT TAC TTC CTT
    Arg Val Pro Lys His Thr Pro Thr Gly Trp Cys Ser Trp Tyr His Tyr Phe Leu
  
```

Figure 10

18/33

*Thermotoga maritima* Alpha-galactosidase  
Complete Gene Sequence (2 of 3)

603	612	621	630	639	648
GAT CTC ACC TGG GAA GAG ACC CTC AAG AAC CTC AAG CTC GCG AAG AAT TTC CCG					
Asp Leu Thr Trp Glu Glu Thr Leu Lys Asn Leu Lys Leu Ala Lys Asn Phe Pro					
657	666	675	684	693	702
TTC GAG GTC TTC CAG ATA GAC GAC GCC TAC GAA AAG CAC ATA GGT GAC TGG CTC					
Phe Glu Val Phe Gln Ile Asp Asp Ala Tyr Glu Lys Asp Ile Gly Asp Trp Leu					
711	720	729	738	747	756
GTG ACA AGA GGA GAC TTT CCA TCG GTG GAA GAG ATG GCA AAA GTT ATA GCG GAA					
Val Thr Arg Gly Asp Phe Pro Ser Val Glu Glu Met Ala Lys Val Ile Ala Glu					
765	774	783	792	801	810
AAC GGT TTC ATC CCG GGC ATA TGG ACC GCC CCG TTC AGT GTT TCT GAA ACC TCG					
Asn Gly Phe Ile Pro Gly Ile Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser					
819	828	837	846	855	864
GAT GTA TTC AAC GAA CAT CCG GAC TGG GTA GTG AAG GAA AAC GGA GAG CCG AAG					
Asp Val Phe Asn Glu His Pro Asp Trp Val Val Lys Glu Asn Gly Glu Pro Lys					
873	882	891	900	909	918
ATG GCT TAC AGA AAC TCG AAC AAA AAG ATA TAC GCC CTC GAT CTT TCG AAA GAT					
Met Ala Tyr Arg Asn Trp Asn Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp					
927	936	945	954	963	972
GAG GTT CTG AAC TGG CTT TTC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC					
Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Arg Lys Met Gly Tyr					
981	990	999	1008	1017	1026
AGG TAC TTC AAG ATC GAC TTT CTC TTC GCG GGT GCC GTT CCA GGA GAA AGA AAA					
Arg Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys					
1035	1044	1053	1062	1071	1080
AAG AAC ATA ACA CCA ATT CAG GCG TTC AGA AAA GGC ATT GAG ACG ATC ACA AAA					
Lys Asn Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys					
1089	1098	1107	1116	1125	1134
GCG GTG GGA GAA GAT TCT TTC ATC CTC GGA TGC GGC TCT CCC CTT CTT CCC GCA					
Ala Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala					
1143	1152	1161	1170	1179	1188
GTC GGA TGC GTC GAC GCG ATG AAG ATA GGA CCT GAC ACT GCG CCG TTC TCG GGA					
Val Gly Cys Val Asp Gly Met Arg Ile Gly Pro Asp Thr Ala Pro Phe Trp Gly					

Figure 10 (Continued)

19/33

*Thermotoga maritima* Alpha-galactosidase  
Complete Gene Sequence (5215)

1197	1206	1215	1224	1233	1242
GAA CAT ATA GAA GAC AAC CCA GCT CCC GCT GCA AGA TOG GCG CTG AGA AAC GCC					
---	---	---	---	---	---
Glu His Ile Glu Asp Asn Gly Ala Pro Ala Ala Arg Trp Ala Leu Arg Asn Ala					
1251	1260	1269	1278	1287	1296
ATA ACG AGG TAC TTC ATG CAC GAC AGG TTC TGG CTG AAC GAC CCC GAC TGT CTG					
---	---	---	---	---	---
Ile Thr Arg Tyr Phe Met His Asp Arg Phe Trp Leu Asn Asp Pro Asp Cys Leu					
1305	1314	1323	1332	1341	1350
ATA CTG AGA GAG GAG AAA ACG GAT CTC ACA CAG AAG GAA AAG GAG CTC TAC TCG					
---	---	---	---	---	---
Ile Leu Arg Glu Glu Lys Thr Asp Leu Thr Gln Lys Glu Lys Glu Leu Tyr Ser					
1359	1368	1377	1386	1395	1404
TAC ACG TGT GGA GTG CTC GAC AAC ATG ATC ATA GAA AGC GAT GAT CTC TCG CTC					
---	---	---	---	---	---
Tyr Thr Cys Gly Val Leu Asp Asn Met Ile Ile Glu Ser Asp Asp Leu Ser Leu					
1413	1422	1431	1440	1449	1458
GTC AGA GAT CAT GCA AAA AAG GTT CTG AAA GAA ACG CTC GAA CTC CTC GGT GGA					
---	---	---	---	---	---
Val Arg Asp His Gly Lys Lys Val Leu Lys Glu Thr Leu Glu Leu Leu Gly Gly					
1467	1476	1485	1494	1503	1512
AGA CCA CGG GTT CAA AAC ATC ATG TCG GAG GAT CTG AGA TAC GAG ATC GTC TCG					
---	---	---	---	---	---
Arg Pro Arg Val Gln Asn Ile Met Ser Glu Asp Leu Arg Tyr Glu Ile Val Ser					
1521	1530	1539	1548	1557	1566
TCT GGC ACT CTC TCA GCA AAC GTC AAG ATC GTG GTC GAT CTG AAC AGC AGA GAG					
---	---	---	---	---	---
Ser Gly Thr Leu Ser Gly Asn Val Lys Ile Val Val Asp Leu Asn Ser Arg Glu					
1575	1584	1593	1602	1611	1620
TAC CAC CTG GAA AAA GAA GGA AAG TCC TCC CTG AAA AAA AGA GTC GTC AAA AGA					
---	---	---	---	---	---
Tyr His Leu Glu Lys Glu Gly Lys Ser Ser Leu Lys Lys Arg Val Val Lys Arg					
1629	1638	1647	1656	1665	
GAA GAC GGA AGA AAC TTC TAC TTC TAC GAA GAG GGT GAG AGA GAA TGA 3'					
---	---	---	---	---	
Glu Asp Gly Arg Asn Phe Tyr Phe Tyr Glu Glu Gly Glu Arg Glu ***					

Figure 10 (Continued)

*Thermotoga maritima*  $\beta$ -mannanase (6GP2)

```

      9      18      27      36      45      54
5' ATG GGG ATT GGT GGC GAC GAC TCC TGG AGC CCG TCA GTA TCG GCG GAA TTC CTT
   ---
Met Gly Ile Gly Gly Asp Asp Ser Trp Ser Pro Ser Val Ser Ala Glu Phe Leu

      63      72      81      90      99      108
TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC GTG AAA
   ---
Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe Val Lys

     117     126     135     144     153     162
GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC ATT GGA AGC
   ---
Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe Ile Gly Ser

     171     180     189     198     207     216
AAC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC AGT GTT CTG GAG
   ---
Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp Ser Val Leu Glu

     225     234     243     252     261     270
AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG GGT TTC CTC GAC GCG
   ---
Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp Gly Phe Leu Asp Gly

     279     288     297     306     315     324
GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT CCT GAG CCC GGT GTT TTC
   ---
Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His Pro Glu Pro Gly Val Phe

     333     342     351     360     369     378
GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC GGT TTC GAA AGA CTC GAC TAC
   ---
Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser Gly Phe Glu Arg Leu Asp Tyr

     387     396     405     414     423     432
ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA AAA CTT GTC ATT GTT CTT GTG AAC
   ---
Thr Val Ala Lys Ala Lys Glu Leu Gly Ile Lys Leu Val Ile Val Leu Val Asn

     441     450     459     468     477     486
AAC TGG CAC GAC TTC GGT GGA ATG AAC CAG TAC GTG AGG TGG TTT GGA GGA ACC
   ---
Asn Trp Asp Asp Phe Gly Gly Met Asn Gln Tyr Val Arg Trp Phe Gly Gly Thr

     495     504     513     522     531     540
CAT CAC GAC GAT TTC TAC AGA GAT GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC
   ---
His His Asp Asp Phe Tyr Arg Asp Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr

```

Figure 11

*Thermotoga maritima*  $\beta$ -mannanase (2000) (continued) (GGP2)

549	558	567	576	585	594
GTC TCC TTT CTC GTA AAC CAT GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA					
Val Ser Phe Leu Val Asn His Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu					
603	612	621	630	639	648
GAG CCC ACC ATC ATG GCC TGG GAG CTT GCA AAC GAA CCG CCG TGT GAG ACG GAC					
Glu Pro Thr Ile Met Ala Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp					
657	666	675	684	693	702
AAA TCG GGG AAC ACG CTC GTT GAG TGG GTG AAG GAG ATG ACC TCC TAC ATA AAG					
Lys Ser Gly Asn Thr Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys					
711	720	729	738	747	756
AGT CTG GAT CCC AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC					
Ser Leu Asp Pro Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn					
765	774	783	792	801	810
TAC GAA GGA TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG					
Tyr Glu Gly Phe Lys Pro Tyr Gly Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp					
819	828	837	846	855	864
TCC GGT GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG					
Ser Gly Val Asp Trp Lys Lys Leu Leu Ser Ile Glu Thr Val Asp Phe Gly Thr					
873	882	891	900	909	918
TTC CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG					
Phe His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp					
927	936	945	954	963	972
GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA CCC					
Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys Pro					
981	990	999	1008	1017	1026
GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA ACG GCC					
Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg Thr Ala					
1035	1044	1053	1062	1071	1080
ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT GGA GCG ATG					
Ile Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp Gly Ala Met					

Figure 11 (Continued)

*Thermotoga maritima*  $\beta$ -mannanase (5827) (continued) (66P2)

1089	1098	1107	1116	1125	1134
TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC GAG AGA GGG TAC					
---	---	---	---	---	---
Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp Glu Arg Gly Tyr					
1143	1152	1161	1170	1179	1188
TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC AGT CCA GAA GCG GAA					
---	---	---	---	---	---
Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp Ser Pro Glu Ala Glu					
1197	1206	1215	1224	1233	1242
CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT GAA GAC ATA AGA GAA GAC					
---	---	---	---	---	---
Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly Glu Asp Ile Arg Glu Asp					
1251	1260	1269	1278	1287	1296
ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG GAG ATC AAA AAG ACC GTG GAA					
---	---	---	---	---	---
Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met Glu Ile Lys Lys Thr Val Glu					
1305	1314	1323	1332	1341	1350
GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC ACG TTT GAA AAG TTG TCT GTC AAA					
---	---	---	---	---	---
Val Arg Ala Gly Val Phe Asp Tyr Ser Asn Thr Phe Glu Lys Leu Ser Val Lys					
1359	1368	1377	1386	1395	1404
GTC GAA GAT CTG GTT TTT GAA AAT GAG ATA GAG CAT CTC GGA TAC GGA ATT TAC					
---	---	---	---	---	---
Val Glu Asp Leu Val Phe Glu Asn Glu Ile Glu His Leu Gly Tyr Gly Ile Tyr					
1413	1422	1431	1440	1449	1458
GGC TTT GAT CTC GAC ACA ACC CGG ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT					
---	---	---	---	---	---
Gly Phe Asp Leu Asp Thr Thr Arg Ile Pro Asp Gly Glu His Glu Met Phe Leu					
1467	1476	1485	1494	1503	1512
GAA GGC CAC TTT CAG GGA AAA ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG					
---	---	---	---	---	---
Glu Gly His Phe Gln Gly Lys Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val					
1521	1530	1539	1548	1557	1566
AAC GAA GCA CCG TAC GTG CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG					
---	---	---	---	---	---
Asn Glu Ala Arg Tyr Val Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu					
1575	1584	1593	1602	1611	1620
GTG AAA AAC TGG TGG AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC					
---	---	---	---	---	---
Val Lys Asn Trp Trp Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp					

Figure 11 (Continued)

*Thermotoga maritima*  $\beta$ -mannanase (~~GAG~~) (continued) (6612)

1629	1638	1647	1656	1665	1674
ATT GAA TGG AAC GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTC					
---	---	---	---	---	---
Ile Glu Trp Asn Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu					
1683	1692	1701	1710	1719	1728
CCC GGA AAG AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC					
---	---	---	---	---	---
Pro Gly Lys Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu					
1737	1746	1755	1764	1773	1782
TCA GAA TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC					
---	---	---	---	---	---
Ser Glu Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu					
1791	1800	1809	1818	1827	1836
AAG GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC					
---	---	---	---	---	---
Lys Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly					
1845	1854	1863	1872	1881	1890
CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC GGA					
---	---	---	---	---	---
Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly Gly					
1899	1908	1917	1926	1935	1944
AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG GGG GTG					
---	---	---	---	---	---
Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala Gly Val					
1953	1962	1971	1980	1989	1998
AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT GGA CCG ATT					
---	---	---	---	---	---
Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp Gly Pro Ile					
2007	2016	2025	2034	2043	
TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG TGA 3'					
---	---	---	---	---	
Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met ***					

Figure 11 (Continued)

AKFII 1a  $\beta$ -mannosidase (63GB1)

5'                    9                    18                    27                    36                    45                    54  
 ATG CTA CCA GAA GAG TTC CTA TGG GGC GTT GCG CAG TCA GGC TTT CAG TTC GAA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe Glu

                  63                    72                    81                    90                    99                    108  
 ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG AAG TGG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Met Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp Lys Trp

                  117                    126                    135                    144                    153                    162  
 GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC CTT CCC GAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp Leu Pro Glu

                  171                    180                    189                    198                    207                    216  
 GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG CTC GCT AAA GGC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Asp Gly Ile Asn Asn Tyr Glu Leu Phe Glu Asn Asp His Lys Leu Ala Lys Gly

                  225                    234                    243                    252                    261                    270  
 CTT GGA CTC AAC GCA TAC AGG ATT GGA ATA GAG TGG AGC AGA ATC TTT CCC TGG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Leu Gly Leu Asn Ala Tyr Arg Ile Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp

                  279                    288                    297                    306                    315                    324  
 CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC ACT TAC GGT TTA GTA AAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp Thr Tyr Gly Leu Val Lys

                  333                    342                    351                    360                    369                    378  
 GAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA CTC GAC AGG CTG GCC AAC AAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu Leu Asp Arg Leu Ala Asn Lys

                  387                    396                    405                    414                    423                    432  
 GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT CAG CAT TTG AGG GAG CTC GGC TTC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Glu Glu Val Met Tyr Tyr Arg Arg Val Ile Gln His Leu Arg Glu Leu Gly Phe

                  441                    450                    459                    468                    477                    486  
 AAG GTC TTC GTT AAC CTC AAC CAC TTC ACG CTT CCA ATA TGG CTC CAC GAC CCG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Lys Val Phe Val Asn Leu Asn His Phe Thr Leu Pro Ile Trp Leu His Asp Pro

                  495                    504                    513                    522                    531                    540  
 ATA GTG GCA AGG GAG AAG GCC CTC ACA AAC GAC AGA ATC GGC TGG GTC TCC CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Ile Val Ala Arg Glu Lys Ala Leu Thr Asn Asp Arg Ile Gly Trp Val Ser Gln

Figure 12



AKPII 1a  $\beta$ -mannosidase (630B1) (continued)

549	558	567	576	585	594
AGG ACA GTT GTT GAG TTT GCC AAG TAT CCT GCT TAC ATC GCC CAT GCG CTC GGA					
Arg Thr Val Val Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala His Ala Leu Gly					
603	612	621	630	639	648
GAC CTC GTG GAC ACA TGG AGC ACC TTC AAC GAA CCT ATG GTA GTT GTG GAG CTC					
Asp Leu Val Asp Thr Trp Ser Thr Phe Asn Glu Pro Met Val Val Val Glu Leu					
657	666	675	684	693	702
GGC TAC CTC GCC CCC TAC TCA GGA TTT CCC CCG GGA GTC ATG AAC CCC GAG GCC					
Gly Tyr Leu Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Met Asn Pro Glu Ala					
711	720	729	738	747	756
GCG AAG CTG GCG ATC CTC AAC ATG ATA AAC GCC CAC GCC TTG GCA TAT AAG ATG					
Ala Lys Leu Ala Ile Leu Asn Met Ile Asn Ala His Ala Leu Ala Tyr Lys Met					
765	774	783	792	801	810
ATA AAG AGG TTC GAC ACC AAG AAG GCC GAT GAG GAT AGC AAG TCC CCT GCG GAC					
Ile Lys Arg Phe Asp Thr Lys Lys Ala Asp Glu Asp Ser Lys Ser Pro Ala Asp					
819	828	837	846	855	864
GTT GCC ATA ATT TAC AAC AAC ATC GGT GTT GCC TAC CCT AAA GAC CCT AAC GAT					
Val Gly Ile Ile Tyr Asn Asn Ile Gly Val Ala Tyr Pro Lys Asp Pro Asn Asp					
873	882	891	900	909	918
CCC AAG GAC GTT AAA GCA GCC GAA AAC GAC AAC TAC TTC CAC AGC GGA CTG TTC					
Pro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn Tyr Phe His Ser Gly Leu Phe					
927	936	945	954	963	972
TTT GAT GCC ATC CAC AAG GGT AAG CTC AAC ATA GAG TTC GAC GGC GAA AAC TTT					
Phe Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe Asp Gly Glu Asn Phe					
981	990	999	1008	1017	1026
GTA AAA GTT AGA CAC CTA AAA GGC AAT GAC TGG ATA GGC CTC AAC TAC TAC ACC					
Val Lys Val Arg His Leu Lys Gly Asn Asp Trp Ile Gly Leu Asn Tyr Tyr Thr					
1035	1044	1053	1062	1071	1080
CGC GAG GTT GTT AGA TAT TCG GAG CCC AAG TTC CCA AGT ATA CCC CTC ATA TCC					
Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro Leu Ile Ser					

Figure 12 (Continued)

AFPII 1a  $\beta$ -mannosidase (630B1) (continued)

1089	1098	1107	1116	1125	1134
TTC AAG GGC GTT CCC AAC TAC GGC TAC TCC TGC AGG CCC GGC ACG ACC TCC GCC					
---	---	---	---	---	---
Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Gly Thr Thr Ser Ala					
1143	1152	1161	1170	1179	1188
GAT GGC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT CCC CAG GGA ATC TAC					
---	---	---	---	---	---
Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Gln Gly Ile Tyr					
1197	1206	1215	1224	1233	1242
GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC ACC GAG AAC					
---	---	---	---	---	---
Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro Val Tyr Val Thr Glu Asn					
1251	1260	1269	1278	1287	1296
GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA TAC TAC ATA GTC AGC CAC GTC					
---	---	---	---	---	---
Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro Tyr Tyr Ile Val Ser His Val					
1305	1314	1323	1332	1341	1350
TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA TAC CCC GTA AAA GGC TAC ATG TAC					
---	---	---	---	---	---
Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly Tyr Pro Val Lys Gly Tyr Met Tyr					
1359	1368	1377	1386	1395	1404
TGG GCG CTT ACG GAT AAC TAC GAG TGG GCC CTC GGC TTC AGC ATG AGG TTT GGT					
---	---	---	---	---	---
Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Leu Gly Phe Ser Met Arg Phe Gly					
1413	1422	1431	1440	1449	1458
CTC TAC AAG GTC GAC CTC ATC TCC AAG GAG AGG ATC CCG AGG GAG AGA AGC GTT					
---	---	---	---	---	---
Leu Tyr Lys Val Asp Leu Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg Ser Val					
1467	1476	1485	1494	1503	1512
GAG ATA TAT CGC AGG ATA GTG CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG					
---	---	---	---	---	---
Glu Ile Tyr Arg Arg Ile Val Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu					
1521	1530	1539			
GAG TTC CTG AAG GGT GAG GAG AAA TGA 3'					
---	---	---			
Glu Phe Leu Lys Gly Glu Glu Lys ***					

Figure 12 (Continued)

## OC1/4V Endoglucanase (33GP1)

5'	ATG	GTA	GAA	AGA	CAC	TTC	AGA	TAT	GTT	CTT	ATT	TGC	ACC	CTG	TTT	CTT	GTT	ATG
	Met	Val	Glu	Arg	His	Phe	Arg	Tyr	Val	Leu	Ile	Cys	Thr	Leu	Phe	Leu	Val	Met
	63					72				81			90		99			108
	CTC	CTA	ATC	TCA	TCC	ACT	CAG	TGT	GGA	AAA	AAT	GAA	CCA	AAC	AAA	AGA	GTG	AAT
	Leu	Leu	Ile	Ser	Ser	Thr	Gln	Cys	Gly	Lys	Asn	Glu	Pro	Asn	Lys	Arg	Val	Asn
	117					126				135			144		153			162
	AGC	ATG	GAA	CAG	TCA	GTT	GCT	GAA	AGT	GAT	AGC	AAC	TCA	GCA	TTT	GAA	TAC	AAC
	Ser	Met	Glu	Gln	Ser	Val	Ala	Glu	Ser	Asp	Ser	Asn	Ser	Ala	Phe	Glu	Tyr	Asn
	171					180				189			198		207			216
	AAA	ATG	GTA	GGT	AAA	GGA	GTA	AAT	ATT	GGA	AAT	GCT	TTA	GAA	GCT	CCT	TTC	GAA
	Lys	Met	Val	Gly	Lys	Gly	Val	Asn	Ile	Gly	Asn	Ala	Leu	Glu	Ala	Pro	Phe	Glu
	225					234				243			252		261			270
	GGA	GCT	TGG	GGA	GTA	AGA	ATT	GAG	GAT	GAA	TAT	TTT	GAG	ATA	ATA	AAG	AAA	AGG
	Gly	Ala	Trp	Gly	Val	Arg	Ile	Glu	Asp	Glu	Tyr	Phe	Glu	Ile	Ile	Lys	Lys	Arg
	279					288				297			306		315			324
	GGA	TTT	GAT	TCT	GTT	AGG	ATT	CCC	ATA	AGA	TGG	TCA	GCA	CAT	ATA	TCC	GAA	AAG
	Gly	Phe	Asp	Ser	Val	Arg	Ile	Pro	Ile	Arg	Trp	Ser	Ala	His	Ile	Ser	Glu	Lys
	333					342				351			360		369			378
	CCA	CCA	TAT	GAT	ATT	GAC	AGG	AAT	TTC	CTC	GAA	AGA	GTT	AAC	CAT	GTT	GTC	GAT
	Pro	Pro	Tyr	Asp	Ile	Asp	Arg	Asn	Phe	Leu	Glu	Arg	Val	Asn	His	Val	Val	Asp
	387					396				405			414		423			432
	AGG	GCT	CTT	GAG	AAT	AAT	TTA	ACA	GTA	ATC	ATC	AAT	ACG	CAC	CAT	TTT	GAA	GAA
	Arg	Ala	Leu	Glu	Asn	Asn	Leu	Thr	Val	Ile	Ile	Asn	Thr	His	His	Phe	Glu	Glu
	441					450				459			468		477			486
	CTC	TAT	CAA	GAA	CCG	GAT	AAA	TAC	GGC	GAT	GTT	TTG	GTG	GAA	ATT	TGG	AGA	CAG
	Leu	Tyr	Gln	Glu	Pro	Asp	Lys	Tyr	Gly	Asp	Val	Leu	Val	Glu	Ile	Trp	Arg	Gln
	495					504				513			522		531			540
	ATT	GCA	AAA	TTC	TTT	AAA	GAT	TAC	CCG	GAA	AAT	CTG	TTC	TTT	GAA	ATC	TAC	AAC
	Ile	Ala	Lys	Phe	Phe	Lys	Asp	Tyr	Pro	Glu	Asn	Leu	Phe	Phe	Glu	Ile	Tyr	Asn

Figure 13

OC1/4V Endoglucanase (33GP1) (continued)

549	558	567	576	585	594
GAG CCT GCT CAG AAC TTG ACA GCT GAA AAA TGG AAC GCA CTT TAT CCA AAA GTG					
---	---	---	---	---	---
Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys Val					
603	612	621	630	639	648
CTC AAA GTT ATC AGG GAG AGC AAT CCA ACC CGG ATT GTC ATT ATC GAT GCT CCA					
---	---	---	---	---	---
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val Ile Ile Asp Ala Pro					
657	666	675	684	693	702
AAC TGG GCA CAC TAT AGC GCA GTG AGA AGT CTA AAA TTA GTC AAC GAC AAA CGC					
---	---	---	---	---	---
Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu Val Asn Asp Lys Arg					
711	720	729	738	747	756
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AAA TTC ACA CAT CAG GGT GCC					
---	---	---	---	---	---
Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe Lys Phe Thr His Gln Gly Ala					
765	774	783	792	801	810
GAA TGG GTT AAT CCC ATC CCA CCT GTT AGG GTT AAG TGG AAT GGC GAG GAA TGG					
---	---	---	---	---	---
Glu Trp Val Asn Pro Ile Pro Pro Val Arg Val Lys Trp Asn Gly Glu Glu Trp					
819	828	837	846	855	864
GAA ATT AAC CAA ATC AGA AGT CAT TTC AAA TAC GTG AGT GAC TGG GCA AAG CAA					
---	---	---	---	---	---
Glu Ile Asn Gln Ile Arg Ser His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln					
873	882	891	900	909	918
AAT AAC GTA CCA ATC TTT CTT GGT GAA TTC GGT GCT TAT TCA AAA GCA GAC ATG					
---	---	---	---	---	---
Asn Asn Val Pro Ile Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Met					
927	936	945	954	963	972
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT GGA					
---	---	---	---	---	---
Asp Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly					
981	990	999	1008	1017	1026
TTT TCA TAC GCG TAT TGG GAA TTT TGT GCA GGA TTT GGC ATA TAC GAT AGA TGG					
---	---	---	---	---	---
Phe Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg Trp					
1035	1044	1053	1062	1071	1080
TCT CAA AAC TGG ATC GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG					
---	---	---	---	---	---
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val Gly Thr Gly Lys Glu					
TAA 3'					
---					
---					

Figure 13 (Continued)

*Thermotoga maritima* Pullulanase (60P3)

```

      9      18      27      36      45      54
5'  ATG GAT CTT ACA AAG GTG GGG ATC ATA GTG AGG CTG AAC GAG TGG CAG GCA AAA
    ---
    Met Asp Leu Thr Lys Val Gly Ile Ile Val Arg Leu Asn Glu Trp Gln Ala Lys

      63      72      81      90      99      108
    GAC GTG GCA AAA GAC AGG TTC ATA GAG ATA AAA GAC GGA AAG GCT GAA GTG TGG
    ---
    Asp Val Ala Lys Asp Arg Phe Ile Glu Ile Lys Asp Gly Lys Ala Glu Val Trp

      117      126      135      144      153      162
    ATA CTC CAG GGA GTG GAA GAG ATT TTC TAC GAA AAA CCA GAC ACA TCT CCC AGA
    ---
    Ile Leu Gln Gly Val Glu Glu Ile Phe Tyr Glu Lys Pro Asp Thr Ser Pro Arg

      171      180      189      198      207      216
    ATC TTC TTC GCA CAG GCA AGG TCG AAC AAG GTG ATC GAG GCT TTT CTG ACC AAT
    ---
    Ile Phe Phe Ala Gln Ala Arg Ser Asn Lys Val Ile Glu Ala Phe Leu Thr Asn

      225      234      243      252      261      270
    CCT GTG GAT ACG AAA AAG AAA GAA CTC TTC AAG GTT ACT GTT GAC GGA AAA GAG
    ---
    Pro Val Asp Thr Lys Lys Lys Glu Leu Phe Lys Val Thr Val Asp Gly Lys Glu

      279      288      297      306      315      324
    ATT CCC GTC TCA AGA GTG GAA AAG GCC GAT CCC ACG GAC ATA GAC GTG ACG AAC
    ---
    Ile Pro Val Ser Arg Val Glu Lys Ala Asp Pro Thr Asp Ile Asp Val Thr Asn

      333      342      351      360      369      378
    TAC GTG AGA ATC GTC CTT TCT GAA TCC CTG AAA GAA GAA GAC CTC AGA AAA GAC
    ---
    Tyr Val Arg Ile Val Leu Ser Glu Ser Leu Lys Glu Glu Asp Leu Arg Lys Asp

      387      396      405      414      423      432
    GTG GAA CTG ATC ATA GAA GGT TAC AAA CCG GCA AGA GTC ATC ATG ATG GAG ATC
    ---
    Val Glu Leu Ile Ile Glu Gly Tyr Lys Pro Ala Arg Val Ile Met Met Glu Ile

      441      450      459      468      477      486
    CTG GAC GAC TAC TAT TAC GAT GGA GAG CTC GGA GCC GTA TAT TCT CCA GAG AAG
    ---
    Leu Asp Asp Tyr Tyr Tyr Asp Gly Glu Leu Gly Ala Val Tyr Ser Pro Glu Lys

      495      504      513      522      531      540
    ACG ATA TTC AGA GTC TGG TCC CCC GTT TCT AAG TGG GTA AAG GTG CTT CTC TTC
    ---
    Thr Ile Phe Arg Val Trp Ser Pro Val Ser Lys Trp Val Lys Val Leu Leu Phe

```

Figure 14

*Thermotoga maritima* Pullulanase (6GP3) (continued)

549	558	567	576	585	594
AAA AAC GGA GAA GAC	ACA GAA CCG TAC CAG	GTT GTG AAC ATG GAA	TAC AAG GGA		
Lys Asn Gly Glu Asp	Thr Glu Pro Tyr Gln	Val Val Asn Met	Glu Tyr Lys Gly		
603	612	621	630	639	648
AAC GGG GTC TGG GAA	GCG GTT GTT GAA	GGC GAT CTC GAC	GGA GTG TTC TAC	CTC	
Asn Gly Val Trp Glu	Ala Val Val Glu	Gly Asp Leu Asp	Gly Val Phe Tyr	Leu	
657	666	675	684	693	702
TAT CAG CTG GAA AAC	TAC GGA AAG ATC	AGA ACA ACC GTC	GAT CCT TAT TCG	AAA	
Tyr Gln Leu Glu Asn	Tyr Gly Lys Ile Arg	Thr Thr Val Asp	Pro Tyr Ser Lys		
711	720	729	738	747	756
GCG GTT TAC GCA AAC	AAC CAA GAG AGC	GCC GTT GTG AAT	CTT GCC AGG	ACA AAC	
Ala Val Tyr Ala Asn	Asn Gln Glu Ser	Ala Val Val Asn	Leu Ala Arg	Thr Asn	
765	774	783	792	801	810
CCA GAA GGA TGG GAA	AAC GAC AGG GGA	CCG AAA ATC GAA	GGA TAC GAA	GAC GCG	
Pro Glu Gly Trp Glu	Asn Asp Arg Gly	Pro Lys Ile Glu	Gly Tyr Glu	Asp Ala	
819	828	837	846	855	864
ATA ATC TAT GAA ATA	CAC ATA GCG GAC	ATC ACA GGA CTC	GAA AAC TCC	GCG GTA	
Ile Ile Tyr Glu Ile	His Ile Ala Asp	Ile Thr Gly Leu	Glu Asn Ser	Gly Val	
873	882	891	900	909	918
AAA AAC AAA GGC CTC	TAT CTC GGG CTC	ACC GAA GAA AAC	ACG AAA GGA	CCG GCG	
Lys Asn Lys Gly Leu	Tyr Leu Gly Leu	Thr Glu Glu Asn	Thr Lys Gly	Pro Gly	
927	936	945	954	963	972
GGT GTG ACA ACA GGC	CTT TCG CAC CTT	GTG GAA CTC GGT	GTT ACA CAC	GTT CAT	
Gly Val Thr Thr Gly	Leu Ser His Leu	Val Glu Leu Gly	Val Thr His	Val His	
981	990	999	1008	1017	1026
ATA CTT CCT TTC TTT	GAT TTC TAC ACA	GGC GAC GAA CTC	GAT AAA GAT	TTC GAG	
Ile Leu Pro Phe Phe	Asp Phe Tyr Thr	Gly Asp Glu Leu	Asp Lys Asp	Phe Glu	
1035	1044	1053	1062	1071	1080
AAG TAC TAC AAC TGG	GGT TAC GAT CCT	TAC CTG TTC ATG	GTT CCG GAG	GGC AGA	
Lys Tyr Tyr Asn Trp	Gly Tyr Asp Pro	Tyr Leu Phe Met	Val Pro Glu	Gly Arg	

Figure 14 (Continued)

**Thermotoga maritima Pullulanase (60P3) (continued)**

1089	1098	1107	1116	1125	1134
TAC TCA ACC GAT CCC AAA AAC CCA CAC ACG AGA ATC AGA GAA GTC AAA GAA ATG					
---	---	---	---	---	---
Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Glu Val Lys Glu Met					
1143	1152	1161	1170	1179	1188
GTC AAA GCC CTT CAC AAA CAC GGT ATA GGT GTG ATT ATG GAC ATG GTG TTC CCT					
---	---	---	---	---	---
Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Met Asp Met Val Phe Pro					
1197	1206	1215	1224	1233	1242
CAC ACC TAC GGT ATA GGC GAA CTC TCT GCG TTC GAT CAG ACG GTG CCG TAC TAC					
---	---	---	---	---	---
His Thr Tyr Gly Ile Gly Glu Leu Ser Ala Phe Asp Gln Thr Val Pro Tyr Tyr					
1251	1260	1269	1278	1287	1296
TTC TAC AGA ATC GAC AAG ACA GGT GCC TAT TTG AAC GAA AGC GGA TGT GGT AAC					
---	---	---	---	---	---
Phe Tyr Arg Ile Asp Lys Thr Gly Ala Tyr Leu Asn Glu Ser Gly Cys Gly Asn					
1305	1314	1323	1332	1341	1350
GTC ATC GCA AGC GAA AGA CCC ATG ATG AGA AAA TTC ATA GTC GAT ACC GTC ACC					
---	---	---	---	---	---
Val Ile Ala Ser Glu Arg Pro Met Met Arg Lys Phe Ile Val Asp Thr Val Thr					
1359	1368	1377	1386	1395	1404
TAC TGG GTA AAG GAG TAT CAC ATA GAC GGA TTC AGG TTC GAT CAG ATG GGT CTC					
---	---	---	---	---	---
Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe Asp Gln Met Gly Leu					
1413	1422	1431	1440	1449	1458
ATC GAC AAA AAG ACA ATG CTC GAA GTC GAA AGA GCT CTT CAT AAA ATC GAT CCA					
---	---	---	---	---	---
Ile Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ala Leu His Lys Ile Asp Pro					
1467	1476	1485	1494	1503	1512
ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGA TGG GGA GCA CCG ATC AGG TTT					
---	---	---	---	---	---
Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly Trp Gly Ala Pro Ile Arg Phe					
1521	1530	1539	1548	1557	1566
GGA AAG AGC GAT GTC GCC GGC ACA CAC GTG GCA GCT TTC AAC GAT GAG TTC AGA					
---	---	---	---	---	---
Gly Lys Ser Asp Val Ala Gly Thr His Val Ala Ala Phe Asn Asp Glu Phe Arg					
1575	1584	1593	1602	1611	1620
GAC GCA ATA AGG GGT TCC GTG TTC AAC CCG AGC GTC AAG GGA TTC GTC ATG GGA					
---	---	---	---	---	---
Asp Ala Ile Arg Gly Ser Val Phe Asn Pro Ser Val Lys Gly Phe Val Met Gly					

Figure 14 (Continued)

*Thermotoga maritima* Pullulanase (6GP3) (continued)

1629	1638	1647	1656	1665	1674
GGA TAC GGA AAG GAA ACC AAG ATC AAA AGG GGT GTT GTT GGA AGC ATA AAC TAC					
---	---	---	---	---	---
Gly Tyr Gly Lys Glu Thr Lys Ile Lys Arg Gly Val Val Gly Ser Ile Asn Tyr					
1683	1692	1701	1710	1719	1728
GAC GGA AAA CTC ATC AAA AGT TTC GCC CTT GAT CCA GAA GAA ACT ATA AAC TAC					
---	---	---	---	---	---
Asp Gly Lys Leu Ile Lys Ser Phe Ala Leu Asp Pro Glu Glu Thr Ile Asn Tyr					
1737	1746	1755	1764	1773	1782
GCA GCG TGT CAC GAC AAC CAC ACA CTG TGG GAC AAG AAC TAC CTT GCC GCC AAA					
---	---	---	---	---	---
Ala Ala Cys His Asp Asn His Thr Leu Trp Asp Lys Asn Tyr Leu Ala Ala Lys					
1791	1800	1809	1818	1827	1836
GCT GAT AAG AAA AAG GAA TGG ACC GAA GAA GAA CTG AAA AAC GCC CAG AAA CTG					
---	---	---	---	---	---
Ala Asp Lys Lys Lys Glu Trp Thr Glu Glu Glu Leu Lys Asn Ala Gln Lys Leu					
1845	1854	1863	1872	1881	1890
GCT GGT GCG ATA CTT CTC ACT TCT CAA GGT GTT CCT TTC CTC CAC GGA GGG CAG					
---	---	---	---	---	---
Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Pro Phe Leu His Gly Gly Gln					
1899	1908	1917	1926	1935	1944
GAC TTC TGC AGG ACG ACG AAT TTC AAC GAC AAC TCC TAC AAC GCC CCT ATC TCG					
---	---	---	---	---	---
Asp Phe Cys Arg Thr Thr Asn Phe Asn Asp Asn Ser Tyr Asn Ala Pro Ile Ser					
1953	1962	1971	1980	1989	1998
ATA AAC GGC TTC GAT TAC GAA AGA AAA CTT CAG TTC ATA GAC GTG TTC AAT TAC					
---	---	---	---	---	---
Ile Asn Gly Phe Asp Tyr Glu Arg Lys Leu Gln Phe Ile Asp Val Phe Asn Tyr					
2007	2016	2025	2034	2043	2052
CAC AAG GGT CTC ATA AAA CTC AGA AAA GAA CAC CCT GCT TTC AGG CTG AAA AAC					
---	---	---	---	---	---
His Lys Gly Leu Ile Lys Leu Arg Lys Glu His Pro Ala Phe Arg Leu Lys Asn					
2061	2070	2079	2088	2097	2106
GCT GAA GAG ATC AAA AAA CAC CTG GAA TTT CTC CCG GCC GGG AGA AGA ATA GTT					
---	---	---	---	---	---
Ala Glu Glu Ile Lys Lys His Leu Glu Phe Leu Pro Gly Gly Arg Arg Ile Val					
2115	2124	2133	2142	2151	2160
GCG TTC ATG CTT AAA GAC CAC GCA GGT GGT GAT CCC TGG AAA GAC ATC GTG GTG					
---	---	---	---	---	---
Ala Phe Met Leu Lys Asp His Ala Gly Gly Asp Pro Trp Lys Asp Ile Val Val					

Figure 14 (Continued)



**Thermotoga maritima Pullulanase (60P3) (continued)**

2169	2178	2187	2196	2205	2214
ATT TAC AAT GGA AAC TTA GAG AAG ACA ACA TAC AAA CTG CCA GAA GGA AAA TGG					
---	---	---	---	---	---
Ile Tyr Asn Gly Asn Leu Glu Lys Thr Thr Tyr Lys Leu Pro Glu Gly Lys Trp					
2223	2232	2241	2250	2259	2268
AAT GTG GTT GTG AAC AGC CAG AAA GCC GGA ACA GAA GTG ATA GAA ACC GTC GAA					
---	---	---	---	---	---
Asn Val Val Val Asn Ser Gln Lys Ala Gly Thr Glu Val Ile Glu Thr Val Glu					
2277	2286	2295	2304	2313	
GGA ACA ATA GAA CTC GAT CCG CTT TCC GCG TAC GTT CTG TAC AGA GAG TGA 3'					
---	---	---	---	---	---
Gly Thr Ile Glu Leu Asp Pro Leu Ser Ala Tyr Val Leu Tyr Arg Glu ***					

**Figure 14 (Continued)**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/00092

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : 435/201, 252.33; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/201, 252.33; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

aps, caplus, biosis

search terms: glycosidase(s), thermococcus, staphylothermus, pyrococcus

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VOORHORST et al. Characterization of the celB gene coding for $\beta$ -glucosidase from the hyperthermophilic archaeon Pyrococcus furiosus and its expression and site-directed mutation in Escherichia coli. J. Bacteriology. December 1995, Vol. 177, No. 24, pages 7105-7111, especially pages 7105, 7106 and 7108.	1-9
Y	Database CAPLUS on STN, CAS, (Columbus, OH, USA), AN 1996:106914, KENGEN et al. "An extremely thermostable .beta.-glucosidase from the hyperthermophilic archaeon Pyrococcus furiosus; a comparison with other glycosidases." Biocatalysis 1994, Vol. 11, No. 2, pages 79-88. Abstract.	1-9

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance		
"E" earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed	"G"	document member of the same patent family

Date of the actual completion of the international search 29 MARCH 1997	Date of mailing of the international search report 09 JUN 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ELIZABETH SLOBODYANSKY Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)\*

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/00092

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	BAUER et al. Comparison of $\beta$ -glucosidase and $\beta$ -mannosidase from the hyperthermophilic archaeon <i>Pyrococcus furiosus</i> . J. Biol. Chem. 27 September 1996, Vol. 271, No. 39, pages 23749-23755, see entire document.	1-9

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/00092

## A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

C12N 9/26, 1/20; C07H 21/04

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

Group I, claims 1-9, drawn to a DNA, a vector comprising the DNA, a cell transformed with the same and a process for producing a peptide.

Group II, claim 10, drawn to an enzyme.

Group III, claim 11, drawn to a method of use of an enzyme.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: A DNA of Group I and an enzyme of Group II are different compounds with different chemical structures and different utilities and therefore do not share a special technical feature. The method of Group III uses an enzyme and therefore does not share a special technical feature with Group I. PCT Rule 1.475(d) does not provide for the multiple products or methods within a single application and therefore unity of invention is lacking with regard to groups I, II and III.